UNIVERSIDADE DE LISBOA

FACULDADE DE FARMÁCIA

DEPARTAMENTO DE MICROBIOLOGIA E IMUNOLOGIA

MYCOBACTERIOPHAGE MS6: EXPLORING THE INVOLVEMENT OF GP1 ON LYSA EXPORT

Francisco André de Lemos Martins

Dissertação de Mestrado

MESTRADO EM CIÊNCIAS BIOFARMACÊUTICAS

2014

UNIVERSIDADE DE LISBOA

FACULDADE DE FARMÁCIA

DEPARTAMENTO DE MICROBIOLOGIA E IMUNOLOGIA

MYCOBACTERIOPHAGE MS6: EXPLORING THE INVOLVEMENT OF GP1 ON LYSA EXPORT

Francisco André de Lemos Martins

Dissertação de Mestrado orientada pela Prof.ª Doutora Madalena Pimentel

MESTRADO EM CIÊNCIAS BIOFARMACÊUTICAS

2014

Acknowledgments

This master dissertation reflects the support of many people who influenced the work in different ways and to whom I would like to thank.

First of all, I would like to express my gratitude to my supervisor, Prof. Madalena Pimentel, for the continuous support during these last 2 years, for her patience, motivation and immense knowledge. Thank you for sharing your wisdom and for allowing me to learn with your experience. Without your supervision and constant help this dissertation would not have been possible.

I am grateful to CPM-URIA of Faculty of Pharmacy, University of Lisbon for allowing me to use their facilities to develop my work.

I want to thank all my CPM colleagues, Adriano, Diane, Pedro and Sofia for all the great moments that we spent together during this journey and for always providing a good working environment.

I also would like to thank the members of Prof. Carlos São-José Lab for sharing their experience, helpful advices and for having proportioned some of the best moments in the lab while we were going through difficult times.

I want to express my absolute gratitude to my friends and family, for being so supporting and understanding during this period of my life.

A sincere thank you to the friends I have made during my academic years, especially to Carlota, Catarina S., Joana F., Joana M., Mafalda, Rafael, Sofia and Vera. Thank you for your unconditional friendship during the last 6 years and for always being by my side. Thank you so much Rafael for all your patience and constantly motivating words encouraging me to move forward.

A very special acknowledgment to my parents and brother for always support my decisions and for being there for me, whenever I needed.

Finally I also would like to thank Dr. Michael Niederweis for providing the anti-MspA antibody that was essential for the completion of this work.

Abstract

Mycobacteriophage Ms6 is a temperate double-stranded DNA (dsDNA) phage that infects the non-pathogenic *Mycobacterium smegmatis*. Similarly to what happens with all other dsDNA phages studied so far, Ms6 must compromise host cell integrity in order to release its progeny at the end of the lytic cycle. Ms6 lytic operon is organized into five genes. In addition to the endolysin (*lysA*) and holin-like genes (*gp4* and *gp5*), two accessory lysis genes are found, *gp1* and *gp3* (*lysB*), which reflects a novel mechanism of phage-mediated lysis. *lysB* encodes an enzyme with lipolytic activity whereas *gp1* encodes a chaperone-like protein. Gp1 interacts with the N-terminal region of LysA and enables its access to the peptidoglycan layer in a holin-independent manner. However, some aspects concerning Gp1 role in the lytic process are not completely clear. In this work we present data obtained using a recombinant Ms6 carrying *gp1* and *lysA* fused to tag sequences. Subcellular fractionation of *M. smegmatis* infected cells revealed that Gp1 is present on the cell wall and cell membrane fractions, while LysA seems to be restricted to the cell wall. Despite the association of Gp1 with the cell envelope, translational fusions with the *E. coli* alkaline phosphatase gene have shown that Gp1 is not endowed with a signal sequence. These results together with the observation that Gp1 is not able to promote the export of the first 60 amino acids of LysA fused to PhoA' suggest that Gp1 and LysA are exported as a complex. The association between the two proteins may be important to keep LysA inactive until the proper time of lysis. The study of bacteriophages opens new perspectives regarding the treatment of bacterial infections and, in this case, it may also contribute to a better understanding of the diverse mechanisms employed by bacteriophages to lyse their hosts.

Keywords: Mycobacteriophage Ms6; mycobacteria; lysis; Ms6 Gp1; secreted endolysins.

Resumo

Os bacteriófagos, ou fagos, são os vírus que infectam bactérias. Estima-se que os fagos constituem a entidade biológica mais abundante do planeta Terra, desempenhando um papel importante na ecologia e evolução microbianas. Os fagos podem apresentar uma grande variedade de morfologias, no entanto, até à data, a maioria dos fagos descritos apresenta cauda e um genoma de DNA em dupla cadeia (dsDNA). Tal como todos os vírus, os bacteriófagos requerem células hospedeiras para se poderem multiplicar de forma a gerar descendência. De acordo com o seu ciclo de infecção, os bacteriófagos de dsDNA podem ser divididos em virulentos, se realizarem um ciclo lítico, ou temperados, se concretizarem um ciclo lítico ou lisogénico. Durante o ciclo lítico o fago infecta células hospedeiras e multiplica-se, produzindo novas partículas virais no seu interior que vão poder infectar outras células. Para iniciar a infecção o fago deve adsorver à superfície da célula bacteriana através do reconhecimento de receptores específicos presentes no envelope celular e posteriormente injectar o seu genoma na célula hospedeira. Depois da injecção do genoma fágico, este utiliza a maquinaria do hospedeiro de forma a gerar novas partículas virais, que são libertadas durante a lise celular induzida pelo fago. Durante o ciclo lisogénico, o DNA fágico, depois de ser injectado para o interior da célula tal como acontece com os fagos virulentos, é geralmente integrado no genoma do seu hospedeiro, sendo transmitido as células-filhas aquando da divisão celular. Sob determinadas condições o ciclo lítico pode ser induzido e nesse caso o metabolismo do hospedeiro é redireccionado para produzir novas partículas virais. O fim do ciclo lítico culmina com a lise da célula hospedeira para que os viriões recém-sintetizados possam infectar novas células e assim gerar nova descendência fágica. Para que isto aconteça os fagos devem comprometer as estruturas responsáveis pela integridade da célula hospedeira, nomeadamente a parede celular.

Os bacteriófagos de dsDNA, como o fago λ, induzem a lise através da síntese de 2 proteínas essenciais: uma endolisina e uma holina. As endolisinas são enzimas com capacidade de hidrolisar o peptidoglicano, enquanto que as holinas são proteínas membranares de pequenas dimensões que conduzem a uma alteração do potencial de membrana e à formação de lesões na membrana citoplasmática, permitindo o acesso da endolisina ao substrato ou a sua activação. As holinas estão descritas como sendo essenciais para determinar o tempo óptimo da lise, de modo a que a libertação de fagos seja produtiva

para a sobrevivência do fago. O modelo de lise, holina-dependente, usado pelo fago λ foi por muito tempo considerado universal, no entanto estudos mais recentes realizados com outros fagos têm revelado que o transporte das endolisinas pode ser feito de forma independente das holinas, nomeadamente através dos sistemas de secreção bacterianos. As endolisinas cujo transporte para o meio extracitoplasmático é independente da holina geralmente apresentam uma sequência sinal que permite a translocação da proteína, através da membrana citoplasmática, utilizando o sistema Sec do hospedeiro. A primeira descrição de uma endolisina contendo uma sequência sinal teve origem em estudos com o fago fOg44 de *Oenococcus oeni*. Neste caso a endolisina (Lys44) é sintetizada com um péptido sinal que é clivado durante o processo de secreção. Mais recentemente têm sido descritas outras endolisinas, nomeadamente de fagos que infectam bactérias Gram-negativas, que apresentam na sua extremidade N-terminal uma região de carácter hidrofóbico, designada por *Signal-Arrest-Release* (SAR), que permite, da mesma forma, o transporte da endolisina através da membrana celular com o auxílio do sistema Sec. Nestes casos, apesar da holina não apresentar um papel activo no transporte da endolisina, esta apresenta um papel crítico na activação das endolisinas e consequente determinação do tempo de lise ideal.

Os fagos que infectam especificamente micobactérias designam-se micobacteriófagos. Este trabalho debruçou-se sobre o bacteriófago Ms6, um micobacteriófago temperado com um genoma de dsDNA que infecta *Mycobacterium smegmatis*. Tal como todos os outros fagos de dsDNA, o fago Ms6 utiliza a estratégia holina-endolisina para comprometer a integridade celular do seu hospedeiro de forma a libertar a progenia fágica no fim do seu ciclo lítico, no entanto o acesso da endolisina ao peptidoglicano é diferente de todos os modelos descritos até à data. O seu operão lítico está organizado em 5 genes. Para além da endolisina (*lysA*) e das holinas (*gp4* e *gp5*), existem dois genes adicionais, *gp1* e *lysB*, que são reflexo de um novo mecanismo de lise. *lysB* codifica uma enzima com actividade lipolítica, enquanto que o gene *gp1* codifica uma proteína com características semelhantes às das chaperonas moleculares. O produto do gene *gp1*, designado Gp1, interage com os primeiros 60 aminoácidos (aa) da região N-terminal da LysA, auxiliando o acesso desta última ao peptidoglicano de forma independente das holinas. No entanto, alguns aspectos relacionados com o papel da Gp1 no processo de lise não são completamente conhecidos. Nomeadamente não se conhece o mecanismo responsável pela manutenção da endolisina

num estado inactivo até ao momento da lise determinado pelas holinas. Com este trabalho pretendemos determinar a localização celular da Gp1 e da LysA durante uma infecção de forma a compreender melhor papel da Gp1 no processo de lise. Usando a técnica *Bacteriophage Recombineering of Electroporated DNA* (BRED) foi construído um micobacteriófago Ms6 recombinante contendo a extremidade 3' do gene *gp1* fundida com um tag c-Myc e a extremidade 3' do gene *lysA* com um tag de 6 histidinas (His_6). Depois de submeter células de *M. smegmatis* infectadas com o fago Ms6 *gp1*-c-Myc *lysA*-His6 a um protocolo de fraccionamento celular foi possível verificar, por Western-blot, a localização de ambas as proteínas. Apesar da sequência aminoacídica da Gp1 não prever a existência de uma sequência sinal foi possível observar que esta proteína se localiza na parede e membrana celulares. Por outro lado a localização da LysA está restrita à parede celular, o que não é surpreendente uma vez que a endolisina do fago Ms6 possui um domínio de ligação ao peptidoglicano (PGRP) entre os aminoácidos 168 e 312. Para além disso, os resultados mostram que ambas as proteínas começam a ser produzidas antes do tempo de lise e estão ausentes da fracção solúvel. Para verificar a possível existência de um péptido sinal na sequência da Gp1 gerou-se uma estirpe recombinante em que a extremidade 3' da *gp1* está fundida com o gene da fosfatase alcalina sem a sequência sinal (*phoA'*). Fusões com o gene *phoA'* são amplamente usadas para determinar a localização celular de proteínas e a existência de sequências sinal, uma vez que esta enzima só é funcionalmente activa no ambiente oxidativo do periplasma. A ausência de actividade enzimática em meio contendo um substrato cromogénico, bem como em ensaios de quantificação em meio líquido indicam que a Gp1 está desprovida de uma sequência sinal. Por último averiguou-se ainda se a Gp1 tem a capacidade de promover a translocação dos primeiros 60 aa da LysA através da membrana citoplasmática, uma vez que estudos prévios mostram que esta região é essencial para o processo de exportação. Usando a mesma estratégia, construiu-se um plasmídeo em que a sequência que codifica os primeiros 60 aa da LysA está fundida com o gene *phoA'* na presença de *gp1*. Os resultados obtidos indicam que apesar da região N-terminal da LysA ser essencial, esta não é suficiente para promover o transporte da PhoA' para o espaço periplasmático de *M. smegmatis* na presença da Gp1.

Os resultados obtidos com este estudo parecem sugerir que a Gp1 e a LysA são exportadas em conjunto tal como acontece com outras proteínas secretadas pelas micobactérias. Para além disso, a formação do complexo Gp1-LysA parece ser importante

para o processo de translocação. De acordo com estas observações colocamos a hipótese de que a Gp1 poderá estar envolvida na manutenção da LysA num estado inactivo até ao momento de lise, uma vez que ambas as proteínas são exportadas enquanto estão a ser sintetizadas, no entanto são necessários estudos adicionais para confirmar esta hipótese. O estudo dos mecanismos de lise usados pelos bacteriófagos abre novas perspectivas no que diz respeito ao tratamento de infecções bacterianas. Para além disso, o estudo da cassete de lise do micobacteriófago Ms6 contribui para uma melhor compreensão dos diversos mecanismos usados pelos bacteriófagos para lisar os seus hospedeiros e lança novas questões relativamente aos mecanismos de secreção usados pelas micobactérias.

Palavras-chave: Micobacteriófago Ms6; micobactérias; lise; Ms6 Gp1; endolisinas secretadas.

Table of Contents

Abbreviations

aa	amino acid
ATCC	American Type Culture Collection
BCIP	5-bromo-4-chloro-3-indolyl phosphate
bp	Base pair
BRED	Bacteriophage recombineering of electroporated DNA
DNA	Deoxyribonucleic acid
ds	Double-stranded
Fig	Figure
GSP	General secretion pathway
His_6	Hexahistidine
ICTV	International Committee on Taxonomy of Viruses
kan	Kanamycin
kb	Kilobase
LB	Luria-Bertani broth
mAGP	Mycolyl arabinogalactan-peptidoglycan
MOI	Multiplicity of infection
mRNA	Messenger ribonucleic acid
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PGRP	Peptidoglycan recognition protein
PhoA	Alkaline phosphatase
pI	Isoelectric point
pmf	Proton-motive force
***p*NPP**	*p*-nitrophenyl phosphate
RBS	Ribosome binding site
RNA	Ribonucleic acid
SAR	Signal-arrest-release
SDS	Sodium dodecyl sulphate

SP	Signal peptide
ss	Single-stranded
TAT	Twin-Arginine Transporter
TBE	Tris-borate-EDTA
TBS	Tris-buffered saline
TDM	Trehalose dimycolate
TMD	Transmembrane domain
TTS	Type III secretion
tRNA	Transfer ribonucleic acid
TTS	Type III secretion
wt	Wild-type

I. INTRODUCTION

1. Bacteriophages: Classification and Life cycle

Bacteriophages, or simply phages, as commonly designated, are the viruses that infect bacteria. Bacteriophages were discovered twice at the beginning of the 20th century. In 1915, the English bacteriologist Frederick Twort described a transmissible lysis in a "micrococcus" and, in 1917, the Canadian Felix d'Herelle, at the Pasteur Institute in Paris, described the lysis of *Shigella* cultures. Twort abandoned his discovery while D'Herelle devoted the rest of his scientific life to bacteriophages and the phage therapy of infectious diseases (Ackermann, 2003). Bacteriophages occur everywhere in the biosphere and have colonised the most inhospitable habitats, such as volcanic hot springs, being among the most abundant biological entities on Earth. It has been estimated that the total number of phages in the biosphere is on the order of 10^{31} particles (Hendrix, 2003). Consequently, they have a major impact on the ecological balance and dynamics of microbial life (Rodriguez-Valera *et al.*, 2009). At the same time, bacteriophages constitute key players in the evolution of bacteria by shaping their genome through horizontal gene transfer (Canchaya *et al.*, 2003; Brüssow *et al.*, 2004).

There are a variety of different morphological types of bacteriophages and taxonomy is based on their shape and size, as well as on the nature of their nucleic acid. The International Committee on Taxonomy of Viruses (ICTV) currently recognizes one order, 14 families and 37 genera (Ackermann, 2009). Bacteriophages are composed of a protein shell, the capsid, often in the shape of icosahedrons that contains the viral genome. Usually it

comprises dsDNA (double-stranded DNA), but there are small phage groups with ssDNA (single-stranded DNA), ssRNA (single-stranded RNA) or dsRNA (double-stranded RNA) genomes. The great majority of phages carry a more or less complex tail to which a base plate, spikes, or tail fibers can be attached. These structures are involved in recognition and attachment to phage receptors present at the bacterial surface (Ackermann, 2009). Tailed dsDNA phages constitute the order Caudovirales, which includes 3 families according to the morphological features of the tail: *Myoviridae* (contractile tail), *Siphoviridae* (long noncontractile tail) and *Podoviridae* (short tail). The remaining phages are classified into 11 families, separated by profound differences in nucleic acid content and structure (Ackermann, 2009).

Because phages consist of a nucleic acid molecule wrapped in a protective coat, they do not have their own metabolism and depend on a host to replicate. Its genome carries genes that direct the synthesis of more phage and can have variable sizes, depending on the phage complexity. For instance, the small RNA phage MS2 has only four genes while the DNA phage T4 has a larger genome comprising more than 200 genes (Snyder and Champness, 2007). Due to their small size, they are usually detected only by the plaques they form on lawns of susceptible host bacteria. Each type of phage makes plaques on only certain host bacteria, which define its host range. Some are very specific and infect only one species of bacteria, while others have multiple hosts (Snyder and Champness, 2007).

According to their cycle of infection, dsDNA phages can be classified as virulent or temperate. Virulent phages undergo a lytic cycle that ends with lysis and death of the host bacteria cell, whereas temperate phages may follow a lytic or a lysogenic cycle (Weinbauer, 2004). During lytic development (Fig. 1), the phage infects a cell and multiplies, producing more phage that can infect other cells. To start the infection, the phage adsorbs to the bacterial surface, throughout the recognition of specific receptors on the cell envelope and injects its genome into the host cell. Then, transcription of the phage genes begins, usually

by the host RNA polymerase. However, not all the phage genes are transcribed simultaneously. Those transcribed soon after infection are called the early genes and encode mostly enzymes involved in DNA synthesis that will help in DNA replication. The rest of the genes, or the late genes, encoding essentially structural components, are transcribed later. After the production of these elements, the viral genome is encapsidated and the phage particles are accumulated in the cytoplasm of the host cell. Finally, the newly synthesized phages are released after host cell lysis and are ready to infect other cells (Weinbauer, 2004; Snyder and Champness, 2007).

In contrast, some phages are able to maintain a stable relationship with the host cell in which they neither multiply nor are lost from the cell. Such a phage is called a temperate phage and its cycle of infection is shown in Figure 1. In the lysogenic state, the phage DNA either is integrated into the host chromosome or replicates as a plasmid. The phage DNA in the lysogenic state is called a prophage, and the bacterium harbouring a prophage is a lysogen for that phage. The first steps of the lysogenic cycle, such as adsorption and injection of viral DNA, are common to the lytic development. However, in this case, phage DNA is integrated into the host genome and it is spread to the daughter cells during host cell division. Under inducing conditions, the prophage is excised from the bacterial chromosome and it can start a lytic cycle (Weinbauer, 2004; Snyder and Champness, 2007).

The final step of the lytic cycle culminates with the release of the newly assembled virions to the extracellular environment, which is extremely important for the phage "survival". If the cell lyses too early, no or very few phages will have been produced. But in contrast, if it lyses too late, time will be lost and the phage will take too long to spread through a population of bacteria to compete effectively. Therefore, the timing of phage progeny release is crucial to maximize both the burst-size, i.e., the number of virions released, and the opportunity to infect new hosts (São-José *et al.*, 2007).

Most of the tailed dsDNA phages achieve the proper time for lysis by the consecutive use of two lysis proteins – holin and endolysin, whereas phages with simpler and smaller genomes use another strategy, involving the action of only one protein which compromises the synthesis of the peptidoglycan (Young and Wang, 2006). There are still the filamentous phages with ssDNA genomes that do not achieve host lysis in order to complete its cycle of infection. These phages are released from their hosts as part of their phage morphogenetic pathway, by a secretion-related mechanism, which maintains bacterial cell structural integrity (Russel, 1995).

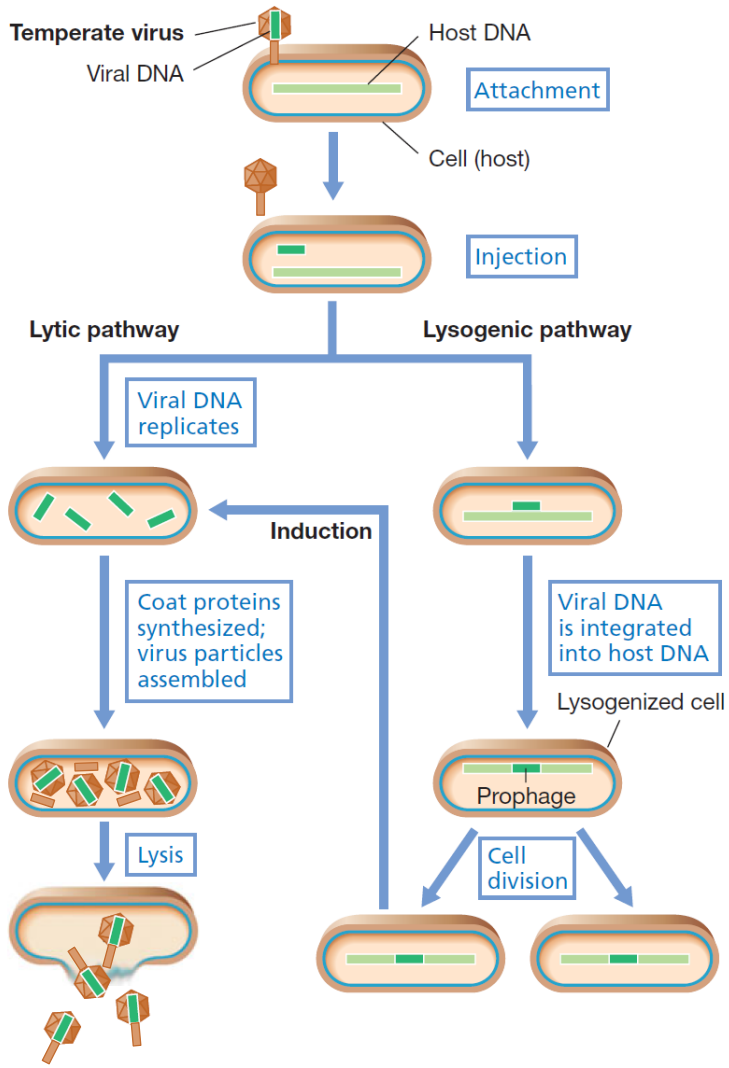


Figure 1. Life cycle of a temperate bacteriophage. The alternatives upon infection are replication and release of mature viruses (lysis) or lysogeny, often by integration of the virus DNA into the host genome, as shown here. The lysogen can be induced to produce mature viruses and lyse. Figure adapted from Madigan *et al.*, (2010).

2. Phage mediated lysis

Most bacteria have a murein cell wall, also known by peptidoglycan, which represents a major challenge to host lysis. Thus, it is fundamental to compromise its integrity to reach the goal of the lytic process. There are two main strategies to accomplish lysis of the host (Young and Wang, 2006). Phages with double-stranded nucleic acid genomes, like phage λ, use the holin-endolisin strategy. The phage elaborates a peptidoglycan hydrolase, an endolysin, specifically dedicated to attack one or more of the three types of peptidoglycan covalent bonds, and a second, membrane-embedded protein, the holin (Young, 2005).

Endolysins can exhibit five major peptidoglycan degrading activities: *N*-acetyl-β-D-glucosaminidases, lytic transglycolases and *N*-acetyl-β-D-muramidases (lysozymes), all hydrolyze the β-1, 4 glycosidic bonds in the murein; *N*-acetylmuramoyl-L-alanine amidases degrade the amide bond connecting the glycan strand to oligopeptide crosslinking chains; and endopeptidases act on the peptide bonds in the same chains (Loessner, 2005; São-José *et al.*, 2007) (Fig. 2).

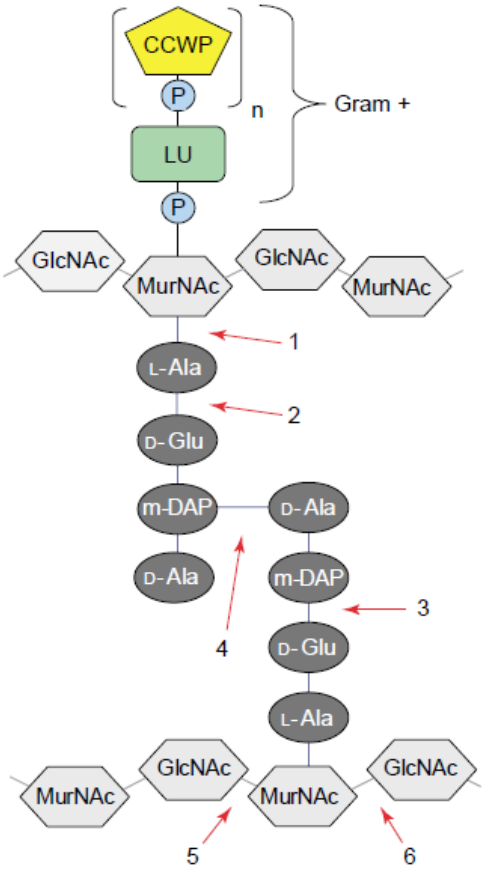


Figure 2. Bacterial cell wall structure and endolysin targets. Detailed structure of the type of peptidoglycan found in *Escherichia coli*. The bonds potentially attacked by endolysins of different enzymatic specificities are indicated by numbers: 1, *N*-acetylmuramoyl-L-alanine amidase; 2, L-alanoyl-D-glutamate endopeptidase; 3, D-glutamyl-m-DAP endopeptidase (this activity has not yet been identified in a phage endolysin); 4, interpeptide bridge-specific endopeptidases; 5, *N*-acetyl-β-D-glucosaminidase; and 6, *N*-acetyl-β-D-muramidase (also known as muramoylhydrolase and "lysozyme") and lytic transglycosylase. Abbreviations: CCWP, carbohydrate cell wall polymer; GlcNAc, *N*-acetyl glucosamine; LU, linkage unit; m-DAP, meso-diaminopimelic acid; MurNAc, *N*-acetyl muramic acid; P, phosphate group. Figure adapted from Loessner (2005).

In general, endolysins from phages of Gram-positive hosts are modular, containing a cell wall binding domain at the C terminus and a catalytic domain at the N-terminus. Moreover, some have two different catalytic domains. The cell wall binding domain directs the enzyme to their substrates and may restrain the enzyme lytic action to a particular type of cell wall (São-José *et al.*, 2007). This binding domain may also prevent the collateral damage of lysis of neighboring cells, since the endolysin is retained in the debris of the lysed host cell and cannot act on other cells (Loessner *et al.*, 2002). With few exceptions described in the literature (Briers *et al.*, 2007; Walmagh *et al.*, 2012), endolysins from phages of Gram-negative hosts are generally small and globular comprising a single domain responsible for the cleavage of a specific peptidoglycan bond (Schmelcher *et al.*, 2012). Recent findings have also shown that some endolysins may be endowed with an N-terminal secretion signal, which targets them to the extra-cytoplasmic media through the host general secretion pathway.

Holins are small, phage-encoded membrane proteins which accumulate in the cytoplasmic membrane of the host. Holins are currently grouped into three classes according to their membrane topology (Fig. 3). The two major classes are class I, with three transmembrane domains (TMDs) (N side out, C side in), and class II, with two TMDs (N side in, C side in). Class III comprises only one gene family and it has only one TMD (N side in, C side out) (São-José *et al.*, 2007). During late gene expression, holins accumulate in the membrane until, at a precise, allele-specific time, a triggering event occurs, resulting in membrane disruption that leads directly, and usually very rapidly, to destruction of the cell wall by the phage-encoded muralytic enzymes. Holins can be prematurely triggered by membrane depolarization with energy poisons such as cyanide and dinitrophenol (Gründling *et al.*, 2001; Young, 2005). This observation together with recent experiments conducted by Young and colleagues (Young and Wang, 2006) led to a model for holin timing. According to these authors, holin molecules accumulate in one or few large two-dimensional aggregates,

or “death rafts”. The rafts maintain the integrity of the membrane until, at some point, a spontaneous aqueous channel develops within the raft. The local collapse of the proton-motive force (pmf) is envisioned to cause the nearby holins to be triggered, just as if an uncoupler had been applied to the cell, and thus the triggered state rapidly propagates throughout the initially triggered raft and other rafts in the cell (Dewey *et al.*, 2010; White *et al.*, 2011).

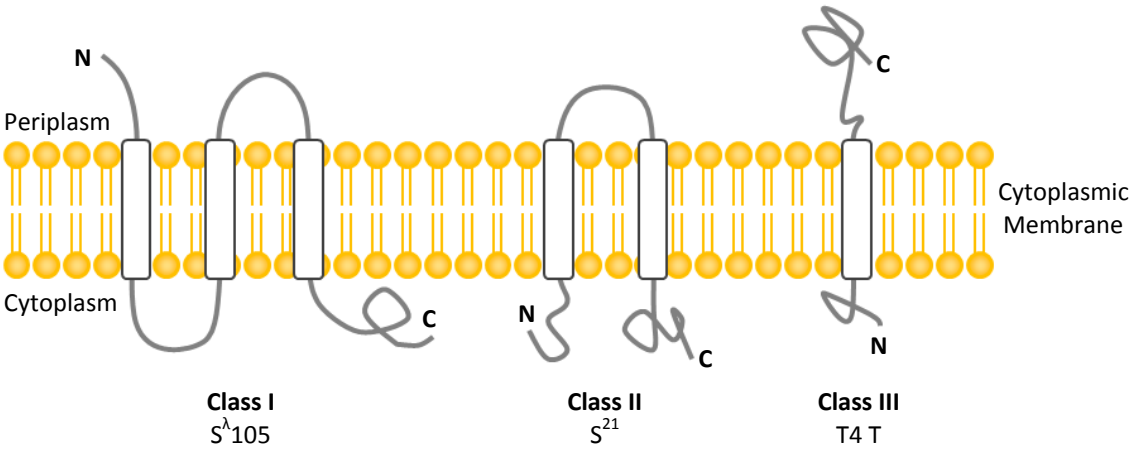


Figure 3. Schematic representation of known topologies of described phage holins. Examples of bacteriophages encoding the different classes of holins are indicated. Figure adapted from São-José *et al.* (2007).

In contrast to the holin-endolysin strategy, icosahedral phages with small, single-stranded genomes, such as phage ϕX174, achieve lysis without encoding a muralytic enzyme. In this case, the phage produces a protein, termed amurin, that causes lysis by acting as a specific inhibitor of an enzyme in the process of murein biosynthesis. This strategy requires actively growing cells and lysis appears to be a consequence of rupture of the cell wall at the developing septum (Bernhardt *et al.*, 2001; Bernhardt *et al.*, 2002).

2.1. The phage λ paradigm

Phage λ is the most well studied phage and its mechanism of lysis was for long considered as a model for most dsDNA phages employing a holin-endolysin lysis strategy. It is a temperate dsDNA phage and infects the Gram-negative *Escherichia coli*. Phage λ lysis

cassette is located immediately downstream of the single late promoter of λ , pR', and is composed of 4 genes translated as a single mRNA, encoding the S_{107} antiholin, the S_{105} holin, the R endolysin and lysis adjuvants Rz and Rz1 (Young, 2002; São-José *et al.*, 2007).

Gene *R* encodes the endolysin which is a 18 kDa (Young and Wang, 2006) murein transglycosylase that accumulates in the cytosol and is suddenly released to the periplasm, at a precise time, through the holes formed by the S holin. The presence of the S_{107} antiholin merely delays lysis onset, allowing for a larger burst-size.

S_{105} holin and S_{107} antiholin are encoded in frame in the same *S* gene and share the same 105 amino acid (aa) sequence, but S_{107} has two extra residues in the N-terminus, Meteonine and Lysine. These extra residues in S_{107} confer two extra positive charges comparing to S_{105}, what results in the altered topology of the antiholin compared to the holin. The S_{105} holin exhibits 3 TMD whereas in the antiholin the first hydrophobic segment is unlikely to span the membrane. It is this difference in topology that confers the phenotype of antiholin and holin. It is worth noting that the dissipation of membrane pmf triggers the translocation of the first TMD of S_{107} which then becomes a topologic homolog of S_{105} with similar hole/lesion-forming properties. The differential expression of S_{107} and S_{105} is due to a Structure Directed Intitiation loop overlapping the Shine-Dalgarno sequence in the *S* mRNA which prevents translation initiation at the S_{107} start codon (Bläsi and Young, 1996).

Genes *Rz* and *Rz1* occupy the distal end of the lysis cassete. *Rz* is a class II inner-membrane protein, whereas *Rz1* gene is embedded in *Rz,* but in a different frame, and encodes an outer-membrane lipoprotein. These two gene products interact with each other and promote the fusion of the inner- and outer-membranes, which facilitates the disruption of the latter (Berry *et al.*, 2008). Both proteins are only essential if the culture medium is supplemented with milimolar concentrations of divalent cations (Zhang and Young, 1999).

Summarizing, the sequence of events during phage λ mediated lysis can be separated in three steps. The first step is the temporally programmed permeabilization of the cytoplasmic

membrane through the formation of micron-scale holes by the holin, which results in the release of the cytoplasmic endolysin (Fig. 4a). The second stage consists on the endolysin-dependent degradation of the murein layer that is followed by the action of Rz and Rz1. The removal of the peptidoglycan meshwork allows lateral, coiled-coil oligomerization of the Rz-Rz1 complexes, which somehow facilitates the disruption of the host outer-membrane, enabling the release of the newly-synthesized virions (Berry *et al.*, 2008; Berry *et al.*, 2010; Berry *et al.*, 2012).

2.2. Sec-mediated Lysis

Despite the fact that phage λ lysis mechanism was considered to be universal, recent studies have shown that endolysins may be transported across the cytoplasmic membrane in a holin-independent manner.

The first indication showing that phage endolysins can exert its functions in a holin-independent way came from the studies on the *Oenococcus oeni* temperate phage fOg44. fOg44 endolysin (Lys44) is synthesized with a typical cleavable signal peptide (SP) that allows its exportation through the bacterial general secretion pathway (GSP) (São-José *et al.*, 2000). Lys44 does not accumulate in the cytoplasm like λ endolysin, but is continuously exported during assembly to the extracytoplasmic environment by the Sec translocon. This raises the important question of how lysis timing is regulated. Surprisingly, all phages that synthesize secreted endolysins or are proposed to do so, seem to encode a holin-like protein as well. If endolysins can use host endogenous pathways to reach the murein cell wall, it would be expected that holins were dispensable. However, in the phage fOg44, a holin function was experimentally demonstrated. Thus, the authors have proposed a model where the activity of the targeted endolysins would be inhibited in the cell wall until dissipation of the membrane potential by the cognate holins. Holins would activate, probably by collapsing the membrane potential, the exported endolysins rather than allowing their release (Fig. 4b).

According to this model, membrane potential is considered a critical parameter in lysis regulation, as seen for the λ lysis mechanism (São-José *et al.*, 2007).

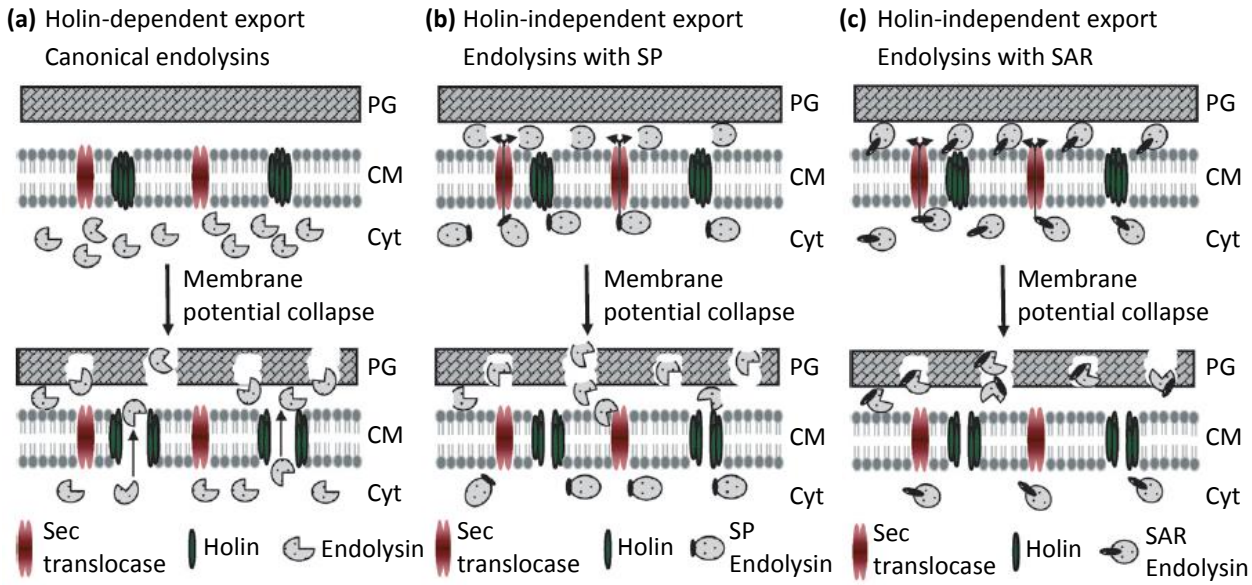


Figure 4. Models for export and activation of phage endolysins. (a) In phages such as λ the export of the active endolysin to the cell wall is done through the holes formed by holins. Holin-independent, Sec-mediated export of endolysins is observed in: (b) phages producing endolysins with typical SP, such as oenophage fOg44; and (c) in phages synthesizing signal-arrest-release (SAR) endolysins, as observed in coliphage P1. When endolysins are exported through the Sec translocase, they are maintained in an inactive state in the cell wall compartment until holins dissipate the membrane pmf. The endolysin activation upon pmf collapse is schematically represented by the change of the enzyme spherical form to a "pacman" shape. Abbreviations: CM, cytoplasmic membrane; Cyt, cytoplasm; PG, peptidoglycan. Figure from Catalão *et al.*, (2013).

More recently, Xu *et al*. (2004) reported the existence of an atypical signal sequence named SAR (signal-arrest-release) in the N-terminal domain of phage P1 endolysin (Lyz^{P1}). Lyz^{P1} export does not require holin action, but is mediated by the N-terminal transmembrane domain and, like fOg44 endolysin, requires host Sec function. However, unlike fOg44, the SAR motif is not proteolytically cleaved. This sequence operates, in a first step, as a signal-arrest domain, directing the endolysin to the periplasm in a membrane-tethered form where it remains enzymatically inactive or restrained from access to the peptidoglycan. In a second step, membrane depolarization triggered by the holin facilitates the instantaneous release of the SAR endolysin from the membrane and its consequent

activation (Xu *et al.*, 2004) (Fig. 4c). Upon release from the membrane, activation of SAR endolysins may be done by an intramolecular thiol-disulfide isomerization involving cysteine residues located at the N-terminal SAR domain, as it happens with P1 Lyz and Lyz^{103} of the *Erwinia amylovora* phage ERA103, which unlocks the enzyme active site (Xu *et al.*, 2005; Kuty *et al.*, 2010). On the other hand, the activation of R^{21}, the endolysin of coliphage 21, does not involve cysteine residues but results from refolding of the SAR domain, which assembles the catalytic triad (Sun *et al.*, 2009).

Sequence comparison has identified additional endolysins with N-terminal hydrophobic sequences in phages infecting Gram-negative hosts. Analysis of these sequences suggests that they could function as a signal anchor and, like Lyz^{P1}, could engage the Sec system (Young and Wang, 2006).

3. Mycobacteriophages

Mycobacteriophages are viruses that specifically infect mycobacterial hosts. The interest in these phages derives in large part from the medical significance and biological idiosyncrasies of their hosts (Hatfull, 2000; Hatfull, 2006). Mycobacteria are acid-fast staining bacteria with characteristic waxy cell walls that can be readily divided into two groups based on their growth rate: slow-growers such as *Mycobacterium tuberculosis* and fast-growers such as *Mycobacterium smegmatis*. Several mycobacterial species are important human and animal pathogens, the most notorious being *M. tuberculosis* and *Mycobacterium leprae*, the causative agents of tuberculosis and leprosy, respectively (Hatfull and Jacobs, 1994; Hatfull, 2006).

Currently, more than 4700 mycobacteriophages have been isolated, most of them having *M. smegmatis* as their host, which leads to the existence of more than 349 complete genome sequences available in GenBank (http://www.phagesdb.org). Based on gross nucleotide sequence similarity, these phages have been grouped into 36 clusters and

subclusters (A-O) and eight singletons that have no close relatives (Hatfull, 2012a; Hatfull, 2012b; Catalão *et al.*, 2013). To date, only mycobacteriophages with a dsDNA genome have been described (Hatfull, 2010; Hatfull *et al.*, 2010; Hatfull, 2012a) and, like all dsDNA phages, they have to face the host cell barriers to release progeny virions at the end of a lytic cycle.

The structure of the mycobacteria cell envelope is much more complex than that of Gram-positive or Gram-negative bacteria. The cytoplasmic membrane, which is structurally and functionally similar to other bacterial cytoplasmic membranes (Daffé *et al.*, 1989), is surrounded by a cell wall core that is composed of peptidoglycan covalently attached to arabinogalactan. This, in turn, is esterified to a mycolic acid layer forming the mycolyl arabinogalactan-peptidoglycan (mAGP) complex (Brennan, 2003). These covalently linked mycolic acids constitute all or part of the inner leaflet of a true outer membrane. The outermost leaflet is composed of various glycolipids, including trehalose mono- and dimycolate, phospholipids and species-specific lipids (Hoffmann *et al.*, 2008; Zuber *et al.*, 2008). Finally, outside of the outer membrane is a layer of proteins, polysaccharides and a small amount of lipids known as the capsule (Lemassu and Daffé, 1994; Lemassu *et al.*, 1996; Sani *et al.*, 2010) (Fig. 5).

Until recently, little was known about the mechanisms underlying mycobacteriophage-induced lysis of mycobacteria, however studies on mycobacteriophage Ms6 provided new insights into the way phages achieve lysis of their hosts. The first report came from the work of Garcia *et al.* (2002), who described the genetic organization of the lysis module of mycobacteriophage Ms6.

3.1. The Lysis Model of Mycobacteriophage Ms6

Mycobacteriophage Ms6 is a temperate phage, isolated from *M. smegmatis* strain HB5688 (Portugal *et al.*, 1989). Ms6 is a dsDNA phage and the length of the genome is over 50 kb with a GC content of 62%. Electronic microscopy studies revealed that phage particles

are composed by an isometric polyhedral head with 80 nm in diameter, hexagonal in shape, and a long non-contractile tail with 210 nm long. These characteristics allowed its classification in the *Siphoviridae* family (Portugal *et al.*, 1989).

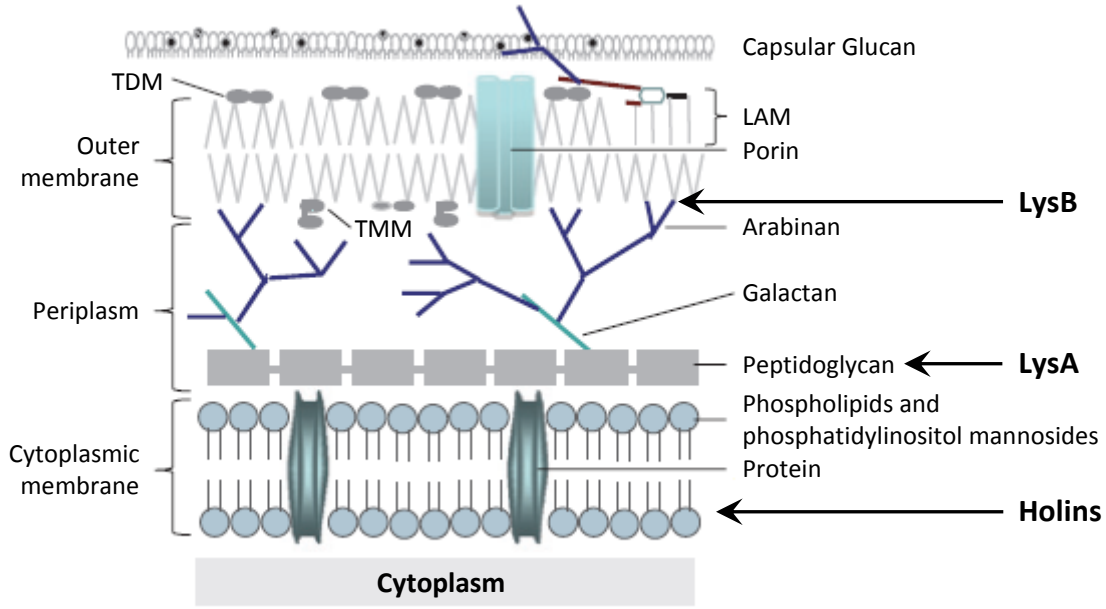


Figure 5. Schematic representation of the mycobacteria cell envelope. The targets of Ms6 lysis proteins are indicated by arrows. LAM, lipoarabinomannan; TDM, trehalose dimycolate; TMM, trehalose monomycolate. Figure adapted from Catalão *et al.* (2013) and Pimentel (2014).

Although the complete analysis of the nucleotide sequence is still not available, some genomic regions are already characterized. The site specific integration locus was identified within a 4.8 kb BglII Ms6 DNA fragment. The integrase gene encodes a protein of 372 aa residues that drives integration into the 3' end of the *M. smegmatis* $tRNA^{Ala}$ gene. The core site, a fragment of 26 bp where the recombination between the phage DNA and the bacterial genome occurs, is positioned near the 5' end of the integrase gene (Freitas-Vieira *et al.*, 1998). 65 bp downstream the integrase gene, and transcribed in the opposite direction, gene *pin* encodes a membrane protein involved in a phage resistance mechanism (Pimentel, 1999). The genetic organization of the lysis module of mycobacteriophage Ms6 was described for the first time in 2002 (Garcia *et al.*, 2002). The Ms6 lysis cassette is composed of five genes clustered downstream of two σ^{70}-like promoters (Fig. 6). This

promoter region (P_{lys}) is separated from the first lysis gene by a leader sequence of 214 bp, in which a transcriptional termination signal was detected, suggesting that an antitermination mechanism is involved in the regulation of Ms6 lysis gene transcription (Garcia *et al.*, 2002).

Figure 6. Genetic organization of the Ms6 lytic operon. The promoter region, P_{lys}, is separated from *gp1* by a leader sequence (L). Direction of the transcription is indicated by an arrow from the P_{lys}. The hairpin shape represents the localization of a transcription termination signal. The beginning of *lysA* overlaps the *gp1* stop codon in a different reading frame. *lysA* encodes two proteins, the full-length $Lysin_{384}$ and the N-terminal truncated version $Lysin_{241}$, which are indicated separately. Figure from Pimentel (2014).

Like all dsDNA phages, Ms6 uses the holin-endolysin strategy to achieve host cell lysis, but the model of lysis is different from all those described so far. In addition to the endolysin (*lysA*) and holin-like genes (*gp4*/*gp5*), two accessory lysis genes were also identified, *gp1* and *gp3* (*lysB*), which led to the development of a peculiar mechanism of endolysin export and specialized functions related to the particular nature of the mycobacteria cell envelope.

Gene *gp1* was identified as encoding a chaperone-like protein that specifically interacts with the N-terminal region of LysA and is involved in its delivery to the peptidoglycan in a holin-independent manner (Catalão *et al.*, 2010). Analysis of the physical and structural properties of Gp1 shows that it shares the properties of molecular chaperones, particularly type III secretion (TTS) system chaperones. *gp1* is positioned immediately upstream of the endolysin gene and overlapped with it, and encodes a small protein of 8.3 kDa with an acidic isoelectric point (pI) of 4.6. Gp1 has the ability to oligomerize and its N-terminal region interacts with the first 60 aa of its effector, LysA (Catalão *et al.*, 2011b). Although not essential for plaque formation, Gp1 is necessary to achieve an efficient lysis, since its absence results in a decrease of approximately 70% in the burst size (Catalão *et al.*, 2010).

The Ms6 endolysin is an enzyme with N-acetylmuramoyl-L-alanine amidase activity, holding a central peptidoglycan recognition protein (PGRP) conserved domain, localized between amino acid residues 168 and 312 (Catalão *et al.*, 2011c). Even though *lysA* was recently shown to encode two proteins designated $Lysin_{384}$ and $Lysin_{241}$ according to the size of the polypetides chain, only $Lysin_{384}$ interacts with Gp1 (Catalão *et al.*, 2011c). $Lysin_{241}$ results from an internal, in-frame second translation initiation site within *lysA* and consequently the N-terminal region that interacts with Gp1 is absent from its amino acid sequence. Despite Ms6 mutants producing only one of the forms of LysA were shown to be viable, both proteins are required for the normal timing, progression and completion of host cell lysis (Catalão *et al.*, 2011c).

Even though Ms6 $Lysin_{384}$ is exported in a holin-independent manner, lysis of *M. smegmatis* infected cells does not occur until holin triggering. Achievement of the correct lysis timing was shown to be dependent on the interaction and concerted action of two membrane proteins with holin-like features, which are encoded by the last two genes comprised in the lytic cassette, *gp4* and *gp5* (Catalão *et al.*, 2011a). *gp4* encodes a small protein of 77 aa, sharing structural characteristics with class II holins. Gene *gp5* is located immediately downstream of *gp4* and produces a 124 aa protein with a predicted TMD at the N-terminus and a highly charged C-terminal, fitting the structural characteristics of class III holins. Ms6 mutated on either of these genes confirms that Gp4 and Gp5 have regulatory roles in determining the timing of lysis. The observation that Gp4 interacts with Gp5 supports the idea that the holin function results from the combined action of Gp4 and Gp5, contributing to the precise adjustment of the timing of hole formation.

As mentioned previously, despite being considered Gram-positive, mycobacteria have a complex cell envelope, including an outer membrane, which imposes an additional challenge to phage release. To overcome this last barrier, Ms6 encodes an enzyme with lipolytic activity, LysB, that cleaves the ester bond between mycolic acids and the arabinogalactan of

the mAGP complex, releasing free mycolic acids (Gil *et al.*, 2008; Gil *et al.*, 2010). Indeed, Ms6 LysB was also shown to hydrolyze other mycobacterial lipid components of the cell envelope, namely the trehalose dimycolate (TDM), a glycolipid involved in the virulence of pathogenic species. Nevertheless, due to the importance of mAGP complex for the stability of the mycobacteria cell envelope it is proposed that the cleavage of the mycobacterial outer membrane from the arabinogalactan-peptidoglycan layer is the primary role of LysB, acting at a later stage of the infection (Gil *et al.*, 2010).

In contrast to what happens with other endolysins exported in a holin-independent manner, mycobacteriophage Ms6 seems to employ a novel and unique mechanism. In this case, the endolysin (LysA) does not possess an N-terminal signal sequence which would allow the transport via the host Sec system. Instead, translocation of Ms6 LysA is assisted by the gene product of *gp1*, a chaperone-like protein. Gp1 specifically binds the Ms6 endolysin and allows the enzyme access to its substrate, the peptidoglycan, by somehow facilitating its secretion across the cytoplasmic membrane. However, how Ms6 endolysin is kept inactive until holin triggering occurs is a question that remains to be elucidated.

4. Objectives

The main goal of this work is to contribute to a better understanding of the role of Gp1 in the export of Ms6 endolysin, LysA, during an infection. Although the organization of the mycobacteriophage Ms6 lytic cassette is well characterized, there are still some missing points concerning the mechanism of lysis, namely the way LysA reaches the peptidoglycan layer and how it is kept inactive until the proper time of lysis.

Does Gp1 have a role in the maintenance of LysA in an inactive state until the correct time of lysis? To provide some insights into this question it is necessary to determine if Gp1 is exported together with LysA. Therefore, with this project we aim to:

- ✓ Detect Gp1 and LysA localization in *M. smegmatis* infected cells, by constructing a recombinant phage where *gp1* and *lysA* are fused to epitope tags;
- ✓ Determine the presence of export signal sequences in Gp1;
- ✓ Test the ability of Gp1 to promote the export of the N-terminal region of LysA.

The study of bacteriophages, with a special focus on its lysis mechanisms, has been growing in the recent years. The emergent problem of antibiotic resistance together with the environmental burden caused by the unrestricted use of antibiotics provides sufficient motivation for developing alternative solutions. Due to its lytic activity, phages could play an important role in treating bacterial infections in humans, animals, aquaculture and crops, as well as in decontaminating food supplies and communal environments (Kutateladze and Adamia, 2010). Even though this study has no direct application in therapeutics or biotechnology, the constantly growing knowledge about bacteriophages lysis systems and its targets on the bacterial cell wall can provide us with some insights into the designing of new therapeutic tools in the future. In addition, in this particular case of secreted endolysins, it may raise new questions regarding mycobacteria protein secretion mechanisms, which still remain poorly understood.

II. MATERIAL AND METHODS

1. Bacterial strains, phages and growth conditions

Bacterial strains, phages and plasmids used in this study are listed in Table 1. *E. coli* was routinely grown at 37ºC in Luria-Bertani (LB) broth with shaking or agar, supplemented with 30 µg mL^{-1} kanamycin (Sigma) or 250 µg mL^{-1} hygromycin (Roche), when appropriated. Unless otherwise indicated, *M. smegmatis* was grown at 37ºC in Middlebrook 7H9 (Difco) with vigorous shaking or in Middlebrook 7H10 (Difco) containing 0.5% glucose and 1 mM $CaCl_2$. When needed, antibiotics were used at the following concentrations: 50 µg mL^{-1} hygromycin and 15 µg mL^{-1} kanamycin.

Phage stocks were obtained by elution of several confluent lysis plates for at least 4 hours at 4ºC with SM buffer (8 mM $MgSO_4$, 100 mM NaCl, 50 mM Tris-HCl, pH 7.5), filter-sterilized and stored at 4ºC. Phage titer was determined by plating serial dilutions of the phage suspension with *M. smegmatis*, as top agar lawns.

2. Preparation and transformation of electrocompetent cells

Preparation of electrocompetent cells of *E. coli* was performed according to Smith *et al.* (1990). Briefly, a culture reaching an optical density at 600 nm (OD_{600}) of 0.6-0.7 in LB was pelleted by centrifugation and washed 3 times with 10% ice-cold Molecular Biology grade Glycerol (AppliChem). After the last centrifugation step, cells were concentrated 250 fold in 10% ice-cold glycerol, frozen and stored at -80ºC.

Induced electrocompetent *M. smegmatis* mc^2155:pJV53 cells were prepared as described previously (van Kessel and Hatfull, 2007). After growth to an OD_{600} of approximately 0.4 in Middlebrook 7H9 (Difco) containing 0.2% succinate and 15 µg mL^{-1} kanamycin, cells were induced with 0.2% acetamide, grown for more 3 hours and then kept on ice for 1 hour and a half. After that period cells were washed three times with ice-cold 10% (v/v) glycerol, concentrated 100-fold and stored at -80ºC. To prepare competent *M. smegmatis* mc^2155 cells, a similar procedure to that described above was used, with the exception of the inducing step.

Electroporation was carried out in a Gene Pulser® Electroporation System (Bio-Rad) using a pulse of 2500 V, 25 µF and 1000 Ω for 0.2 cm cuvettes or 1800 V, 25 µF and 200 Ω for 0.1 cm cuvettes.

3. Phage DNA extraction

Phage DNA extraction was adapted from Sambrook and Russel (2001). A phage stock sample was incubated with 50 µg mL^{-1} of proteinase K (AppliChem) and 0.5% (w/v) SDS for 1 hour at 56ºC to disrupt the phage capsid. The mixture was washed three times with an equal volume of Phenol:Chloroform:Isoamyl alcohol 25:24:1 (AppliChem) and one last time with chloroform. The aqueous phase was recovered and the DNA was precipitated by the addition of 3M sodium acetate and cold isopropanol. After incubation at -20ºC for at least 30 min, the DNA was collected by centrifugation at 4ºC (maximum speed for 45 min), washed with 70% (v/v) ethanol and dried at 37ºC. DNA was resuspended in sterile distilled water and stored at -20ºC.

Table 1. Strains, bacteriophages and plasmids used in this study.

Strains, bacteriophages, plasmids	Description	Reference/ Source
Escherichia coli		
JM109	*recA1 endA1 gur96 thi hsdR17 supE44 relA1* Δ(*lac-proAB*) [F' *traD36 proAB lacI*^{q}*Z*ΔM15]	Stratagene
Mycobacterium smegmatis		
mc^{2}155	High-transformation-efficiency mutant of *M. smegmatis* ATCC607	Snapper *et al.* (1990)
mc^{2}155:pJV53	*M. smegmatis* mc^{2}155 carrying a plasmid expressing recombineering functions; Kan^{R}	van Kessel and Hatfull (2007)
Bacteriophage Ms6		
wt	Temperate bacteriophage from *M. smegmatis*	Portugal *et al.* (1989)
lysA-His$_{6}$	His$_{6}$-tag insertion at the C-terminus of Ms6 LysA	Catalão *et al.* (2010)
gp1-c-Myc *lysA*-His$_{6}$	c-Myc and His$_{6}$-tag insertion at the C-terminus of Ms6 Gp1 and LysA, respectively	This study
Plasmids		
pSMT3[19-phoA]	Mycobacteria plasmid containing the structural gene for *E. coli* PhoA; Hyg^{R}	Herrmann *et al.* (1996)
pVVAP	Mycobacteria shuttle vector carrying the acetamidase promoter; Kan^{R}	V. Visa and M. McNeil, unpublished
pFM1	*gp1-phoA'* fusion cloned in pVVAP	This study
pFM2	*19kDa$_{50aa}$-phoA'* fusion cloned in pVVAP	This study
pFM3	*phoA'* cloned in pVVAP	This study
pFM4	*gp1* and the sequence encoding the first 60 aa of LysA fused to *phoA'* cloned in pVVAP	This study
pFM5	The sequence enconding the first 60 aa of LysA fused to *phoA'* cloned in pVVAP	This study
pFM6	*gp1* and *phoA'* cloned in pVVAP	This study

4. DNA manipulation and purification

DNA fragments were amplified by PCR using the NZYTaq 2x Green Master Mix (NZYTech) or the Pfu DNA Polymerase (Promega) when high fidelity PCR products were required. PCR amplifications were carried out in accordance to the manufacturer instructions, in a standard thermocycler (VWR DOPPIO Thermal Cycler). The amplification products were analysed by electrophoresis in 0.7 to 1.2% (w/v) agarose gels containing GreenSafe Premium

(NZYTech) and using 0.5X Tris-Borate-EDTA (TBE) as electrophoresis buffer (Sambrook and Russel, 2003). Gels were analysed under ultraviolet light and pictures were acquired using ChemiDoc/GelDoc system (BIO-RAD). The size of the fragments was estimated by comparison with DNA ladders (NZYDNA Ladder VI, NZYTech; GeneRuler 100 bp Plus, Fermentas or GeneRuler 1 kb DNA Ladder, Thermo Scientific) run along with DNA samples.

Purification of DNA fragments obtained from PCR amplification or enzymatic restriction was performed using MinElute® PCR Purification Kit (QIAGEN) or Invisorb® Spin DNA Extraction Kit (Invitek). Plasmid DNA purification was carried out using the NZYMiniprep (NZYTech) according to the protocol recommended by the manufacturer. DNA was eluted in a minimal volume of sterile distilled water and stored at -20ºC. Restriction enzymes (FastDigest®, Fermentas) and T4 DNA ligase (New England Biolabs) were used according to supplier's instructions. All oligonucleotides were from Thermo Scientific and are listed in Table 2.

When necessary, DNA concentration was determined by spectrophotometry using NanoDrop® ND-1000 Spectrophotometer (Thermo Scientific). Samples purity was evaluated based on $A_{260/A280}$ ratio.

5. Construction of the recombinant Ms6 *gp1*-c-Myc *lysA*-His_6 phage

Construction of the Ms6 double mutant phage was performed according to the Bacteriophage Recombineering of Electroporated DNA (BRED) technology described by Marinelli *et al.* (2008).

10 ng of a 110-base oligonucleotide designated Prgp1c-Myc, containing the c-Myc tag sequence flanked by 40-base of homology upstream and downstream to the Ms6 *lysA*-His6 genome region to be altered, was extended by PCR using two 75-base primers designated PrExtgp1c-MycFwd and PrExtgp1c-MycRv. These primers overlap 25-base on the ends of the 110-mer oligonucleotide and add additional 50 bp of homology to each end, generating a

210-bp recombineering substrate. The final 210-bp dsDNA PCR product was purified using MinElute® PCR Purification Kit (Qiagen) following manufacturer's instructions.

120.5 ng of the purified recombineering substrate were co-electroporated with 1330 ng of Ms6 *lysA*-His_6 genomic DNA into electrocompetent *M. smegmatis* mc^2155:pJV53 previously induced for recombineering functions using 0.2 cm cuvettes. Cells were resuspended in 7H9 with 0.5% glucose and 1 mM $CaCl_2$, incubated at 37ºC for 2 hours and plated as top agar lawns with *M. smegmatis*. After overnight (ON) incubation at 37ºC phage plaques were picked into 100 µL of SM buffer, eluted for 2 hours at room temperature and analysed by PCR with primers Prc-MyctagFwd and PrlysARv to detect the c-Myc tag insertion. Mixed primary plaques containing both the recombinant and the *wt* phage were eluted as described above and serial dilutions were plated with *M. smegmatis*. Secondary plaques were screened by PCR for mutant detection and re-plated with *M. smegmatis*. This process was repeated until only pure recombinant phages were detected in plaques.

6. One-step growth curves

One-step growth curves were carried out essentially as described previously (Catalão *et al.*, 2010). *M. smegmatis* cells were grown to an OD_{600} of 0.5, pelleted and resuspended in 1 mL of phage suspension (Ms6*wt* or Ms6 *gp1*-c-Myc *lysA*-His_6) supplemented with 1 mM $CaCl_2$, using a multiplicity of infection (MOI) of 1. The mixture was incubated 50 min at 37°C for phage adsorption, and then 100 µL of 0.4% H_2SO_4 were added for 5 min to inactivate non-adsorbed phages. The suspension was neutralized with 100 µL of 0.4% NaOH and diluted 100-fold into fresh pre-warmed 7H9 with 0.5% glucose and 1 mM $CaCl_2$. 1 mL samples were collected every 30 min for a period of 240 min and 100 µL of serial dilutions of each sample were plated with 200 µL of *M. smegmatis* cells on 7H10 supplemented with 0.5% glucose and 1 mM $CaCl_2$, as top agar lawns. Phage titer for each sample was

determined after an ON incubation at 37°C. Results are averages of three independent experiments.

Table 2. Oligonucleotides used in this study.

Primer	Sequence 5' → 3'	Features
Primers used to construct the recombinant Ms6 *gp1*-c-Myc *lysA*-His_6 phage		
Prgp1c-Myc	CTCCATCCCCGTCCTCGGCGGAATCCTCGGGAGCAAACGGGAACAGAAACTGATCAGCGAAGAGGATCTGTGACGGGAGCAAACGGTGACCACGAAAGATCAAGTCGCCC	c-Myc tag sequence[a]
PrExtgp1c-MycFwd	CTGACCAACCTTCCAGCGCAAGTCATGGACATCATCGACAGCGCGCTGCGCTCCATCCCCGTCCTCGGCGGAATC	Overlaps the 5' end of Prgp1c-Myc
PrExtgp1c-MycRv	GCATTCGCTGCGGGTGTAGCCGCGCGCCTTGGCTTCGGCGATGGTGATTTGGGCGACTTGATCTTTCGTGGTCAC	Overlaps the 3' end of Prgp1c-Myc
Prc-MyctagFwd	CTCGGGAGCAAACGGGAACAGAAACTG	Primer specific to c-Myc tag
PrlysARv	GTCGAAGCGGTGTGGGTAGGAGCCG	Flanking primer
Primers used to clone in pVVAP		
PrphoABamHIFwd	GCCGGATCCCCTGTTCTGGAAAACC	BamHI site[a]
PrphoAHis6HindIIIRv	GTAAGCTTTTTCAGCCCCAGAGCGG	HindIII site[a]
PrATGgp1NdeIFwd	GATCATATGGACCGCTTAGGCATCG	NdeI site[a]
Prgp1-SGGGS-BamHIRv	CTAGGATCC**GCTGCCGCCGCCGCT**CCGTTTGCTCCCGAG	BamHI site and SGGGS 3' linker[a,b]
PrATG19kDaNdeIFwd	CTTCATATGAAGCGTGGACTGACGGTC	NdeI site[a]
PrATGphoANdeIFwd	CATCATATGCCTGTTCTGGAAAACCG	NdeI site[a]
PrlysA60aaKpnIRv	CAGGTACCGTGTGGGTAGGAGCCGTCC	KpnI site[a]
PrATGlysANdeIFwd	GATCATATGACCACGAAAGATCAAGTCGC	NdeI site[a]
PrphoAKpnIFwd	CAGGTACCCCTGTTCTGGAAAACCG	KpnI site[a]
Prgp1KpnIRv	CAGGTACCTCACCGTTTGCTCCCGAG	KpnI site[a]
PrRBSphoAKpnIFwd	CAGGTACCAGGAGCACAGGGTGCCTGTTCTGGAAAACC	KpnI site[a]
PrpVVAPFwd	GCAGTTGTTCTCGCATACCCCATC	Flanking primer
PrpVVAPRv	GGCCCAGTCTTTCGACTGAGCCT	Flanking primer

[a]Underlined nucleotides indicate the sequence encoding the c-Myc tag and restriction sites;
[b]Bases in bold show the sequence encoding the SGGGS linker.

7. Gp1-c-Myc and LysA-His$_6$ expression in *M. smegmatis* infected cells and subcellular fractionation

A mid-log phase (OD_{600} between 0.8-1) growing culture of *M. smegmatis* was infected with Ms6 *gp1*-c-Myc *lysA*-His_6 at an approximate MOI of 100 and incubated at 37ºC for 50 minutes. 100 mL samples were collected at 30 min intervals for a period of 2 hours. Cells were washed once in the same volume of ice-cold phosphate-buffered saline (PBS), pelleted by centrifugation and frozen at -20ºC. After thawing, cells were resuspended in 10 mL of PBS containing a cocktail of protease inhibitors (Protease Inhibitor Cocktail Set I, Calbiochem) and lysed using a French press (4 passages at 15000 pounds). Unbroken cells were pelleted at 3000 x *g* for 20 min to generate a clarified whole-cell lysate.

Preparation of crude cell wall, membrane and soluble fractions by differential ultracentrifugation was adapted from Rezwan *et al.* (2007) and Gibbons *et al.* (2007). All the centrifugation steps were performed at 4ºC using an Optima™ L-100 XP ultracentrifuge (Beckman Coulter™). Clarified whole-cell lysates were centrifuged at 50 000 x *g* for 30 min to pellet the cell wall. The resulting supernatants were centrifuged at 100 000 x *g* for 4 hours to separate the membrane fraction from the soluble fraction. Cell wall and membrane fractions were washed once in the same volume of PBS and resuspended in 500 µl of PBS.

Protein concentrations were estimated using the Bio-Rad protein assay and 10 µg of total protein from each fraction were separated by SDS-PAGE (Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis). For the anti-MspA blot, 100 ng of each subcellular fraction were analysed.

8. Protein analysis by SDS-PAGE and Western-blot

Protein extracts were mixed with Laemmli buffer 5x (25% β-mercaptoethanol, 1% Bromophenol blue, 50% Glycerol, 10% SDS, 300 mM Tris-HCl) and denatured at 100ºC for 10 min. The same protein amount or the same volume of protein sample was loaded on a SDS-

PAGE gel and protein separation was performed by electrophoresis at 120 V (Mini-PROTEAN® 3 Cell, BIO-RAD). Total proteins were directly visualized by staining polyacrylamide gels after electrophoresis with BlueSafe (NZYTech), a safer alternative to the traditional Coomassie Blue staining.

For immunoblotting, proteins were transferred to a 0.45 µm pore size nitrocellulose membrane (Amersham™ Hybond C-extra, GE Healthcare) at 250 mA for 1 hour in a wet transfer System (Mini Trans-Blot® Electrophoretic Transfer Cell, BIO-RAD). Membranes were blocked in TBS-T (137 mM NaCl, 20 mM Tris-HCl, 0.1% Tween 20, pH 7.6) containing 5% non-fat dried milk (Molico®, Nestlé) overnight at 4ºC or for 1 hour at room temperature with shaking. Incubation with appropriate antibodies (Table 3) was done for 1 hour at room temperature with shaking in TBS-T containing 1% non-fat dried milk. Membranes were washed 4 to 6 times in TBS-T and, when necessary, incubated with the secondary antibodies for 1 hour at room temperature with shaking and then washed again. Blots were developed using the Amersham™ ECL™ Prime Western blotting detection reagent (GE Healthcare) according to manufacturer's instructions and exposed to X-ray film (Amersham Hyperfil™, GE Healthcare). Protein molecular masses were calculated using the prediction programme Compute pI/Mw from the Expasy Proteomics Server of the Swiss Institute of Bioinformatics (http://www.expasy.org).

Table 3. Antibodies used for immunoblotting**.**

Antibody	**Dilution**	**Supplier**
Anti-His6-Peroxidase	1:2000	Roche
Anti-c-Myc-Peroxidase	1:5000	Roche
Mouse Monoclonal Anti-Mycobacterium tuberculosis KatG (Gene Rv1908c)	1:1000	BEI Resources
Rabbit Polyclonal Anti-MspA antiserum	1:5000	Michael Niederweis
Goat Anti-Mouse IgG Horseradish Peroxidase Conjugate	1:5000	Bio-Rad
Goat Anti-Rabbit IgG Horseradish Peroxidase Conjugate	1:10000	Bio-Rad

9. Plasmid construction

To express translational fusions to the *E. coli* alkaline phosphatase gene (*phoA*) in *M. smegmatis*, DNA fragments were cloned in pVVAP (Fig. 7). pVVAP is a replicating shuttle vector that allows the expression of C-terminally His_6-tagged proteins driven by the acetamidase promoter (Parish *et al.*, 1997). In addition, for optimal expression in mycobacteria, pVVAP contains a consensus ribosomal binding site (RBS) and an ATG start codon which is embedded in NdeI restriction site.

To construct a Gp1–PhoA hybrid protein, pSMT3 [19-phoA] (Herrmann *et al.*, 1996) was used as template to amplify *phoA* gene omitting the PhoA signal peptide (*phoA'*), using primers PrphoABamHIFwd/PrphoAHis6HindIIIRv. *gp1* lacking its start codon was amplified by PCR from Ms6*wt* genomic DNA using the forward primer PrATGgp1NdeIFwd, which includes a NdeI restriction site, and the reverse primer Prgp1-SGGGS-BamHIRv, carrying a SGGGS 3' linker and a BamHI restriction site. *gp1* and *phoA'* PCR products were digested with NdeI/BamHI and BamHI/HindIII, respectively, and ligated in a single ligation reaction with pVVAP previously digested with NdeI/HindIII to generate pFM1. Control plasmids carrying *19kDa$_{50aa}$-phoA'* fusion or *phoA'* alone were constructed directly from pSMT3 [19-phoA] using the pair of primers PrATG19kDaNdeIFwd/PrphoAHis6HindIIIRv and PrATGphoANdeIFwd/PrphoAHis6HindIIIRv, respectively. Both PCR products were double-digested with NdeI/HindIII and inserted into the corresponding sites of pVVAP, generating pFM2 and pFM3, respectively.

DNA fragments containing *gp1* and the sequence encoding the first 60 aa of LysA (*gp1-lysA$_{60aa}$*), or the first 60 aa of LysA alone (*lysA$_{60aa}$*) were PCR amplified from Ms6*wt* DNA using the pair of primers PrATGgp1NdeIFwd/PrlysA60aaKpnIRv or PrATGlysANdeIFwd/PrlysA60aaKpnIRv, respectively. Both fragments were digested with NdeI and KpnI. *phoA'* was obtained by PCR from pSMT3 [19-phoA] with primers PrphoAKpnIFwd/PrphoAHis6HindIIIRv and digested with KpnI and HindIII. *gp1-lysA$_{60aa}$* or

lysA$_{60aa}$ and *phoA'* restriction products were ligated into pVVAP NdeI/HindIII restriction sites, generating pFM4 and pFM5, respectively. Plasmid pFM6 was constructed by ligation of *gp1*, including its stop codon, and *phoA'*, containing translational signals, with pVVAP. *gp1* fragment was amplified with primers PrATGgp1NdeIFwd and Prgp1KpnIRv, which encompasses *gp1* stop codon, and digested with NdeI/KpnI. After generation by PCR with primers PrRBSphoAKpnIFwd, which provides a RBS sequence and a GTG start codon, and PrphoAHis6HindIIIRv, *phoA'* fragment was digested with KpnI/HindIII. Both fragments were cloned into the corresponding NdeI/HindIII sites of pVVAP. Ligation mixtures were inserted into *E. coli* JM109 by electroporation and recombinant clones were selected by colony PCR using primers PrpVVAPFwd/PrpVVAPRv that flank the insertion site. All constructs were verified by DNA sequencing at Macrogen and then introduced into *M. smegmatis* cells by electroporation.

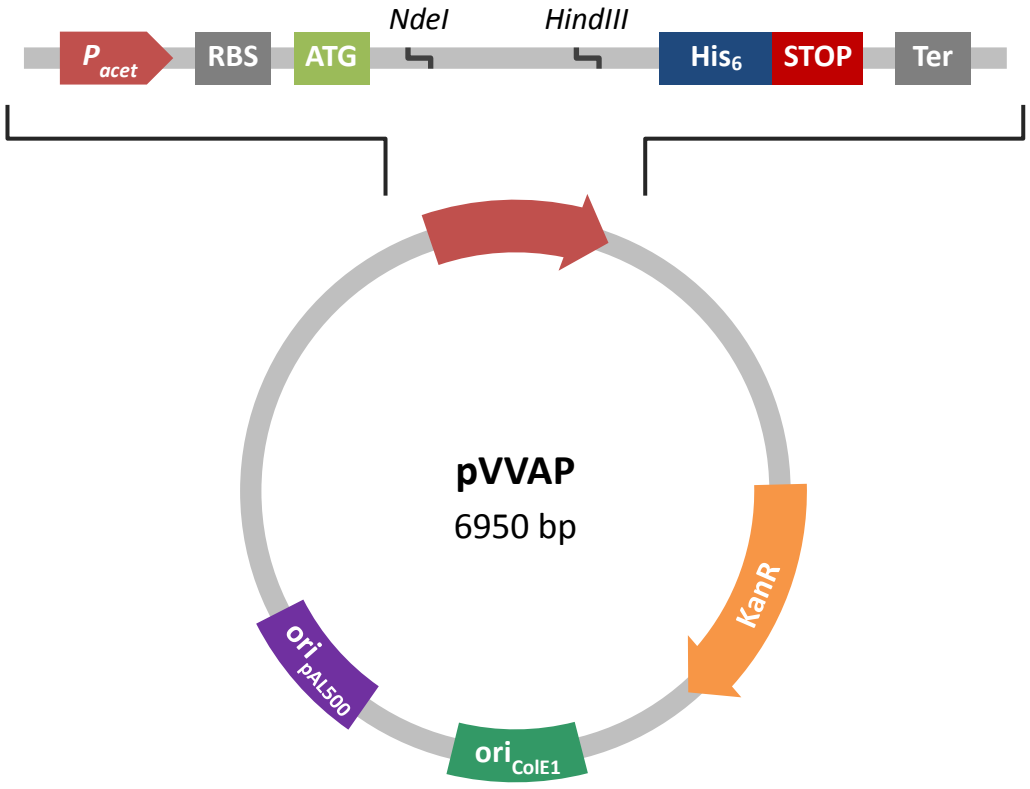


Figure 7. Schematic representation of pVVAP. ATG – start codon; His_6 – sequence encoding a stretch of six histidines; KanR – kanamycin resistance gene; ori_{ColE1} – ColE1 origin of replication; $ori_{pAL5000}$ – pAL5000 origin of replication; P_{acet} – acetamidase promoter; RBS – ribosome binding site; STOP – stop codon; Ter – *rrnB* T1 transcriptional termination region.

10. Detection and quantification of alkaline phosphatase activity

To detect PhoA activity on plates, recombinant strains were streaked on Middlebrook 7H9 agar supplemented with 60 μg mL^{-1} 5-bromo-4-chloro-3-indolyl phosphate (BCIP), 0.2% succinate, 0.2% acetamide and 15 μg mL^{-1} kanamycin. To enhance the visibility of the blue colour on plates malachite green was omitted from the recipe and 7H9 supplemented with agar was used instead of 7H10 (Wolschendorf *et al.*, 2007). Plates were put into a sealed plastic bag to keep them moist and incubated in the dark at 37°C. The colour of the colonies was examined daily for a period of up to 7 days.

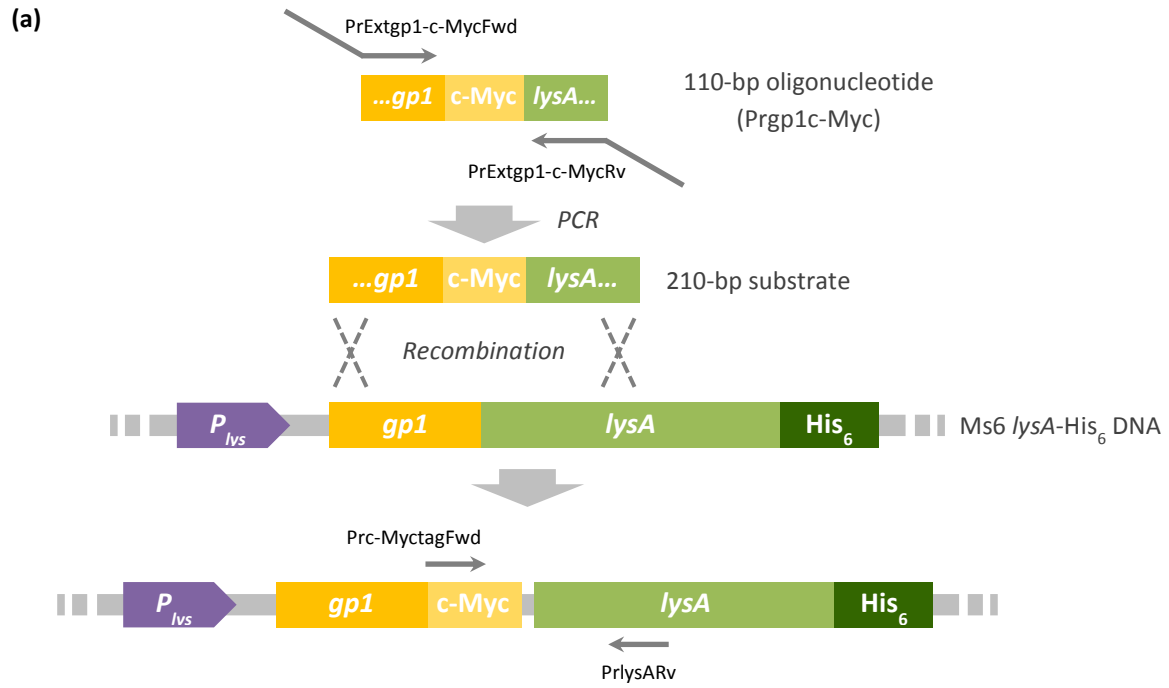
PhoA activity of *M. smegmatis* in liquid cultures was measured as described previously (Catalão *et al.*, 2010). *M. smegmatis* cells were grown in 7H9 supplemented with 0.2% Succinate and 15 μg mL^{-1} kanamycin at 37ºC with shaking. When OD_{600} reached approximately 0.4, cultures were induced with 0.2% acetamide for 6h. Then, cells were harvested by centrifugation at 4ºC, resuspended in the same volume of ice-cold 1 M Tris-HCl, pH 8.0 and the OD_{600} of the cell suspension was determined. Cells were kept on ice during all steps. PhoA activity was measured in whole cells by adding 100 μL of the substrate solution (0.2 M *p*-nitrophenyl phosphate disodium salt hexahydrate (Sigma) dissolved in 1 M Tris-HCl, pH 8.0) to 900 μL of cell suspension. Reactions proceeded at 37ºC until a yellow colour developed and stopped with 100 μL of 1 M K_2HPO_4. Cells were centrifuged at 12 000 x *g* for 2 min and the supernatant read at 405 nm. Enzyme activity was expressed in arbitrary units of OD_{405} mL^{-1} of culture per minute.

III. RESULTS

1. Construction of the recombinant Ms6 *gp1*-c-Myc *lysA*-His_6 phage

Previous studies have shown that Gp1 behaves as a chaperone-like protein and specifically interacts with the N-terminal region of LysA to assist its translocation across the to the extracytoplasmic environment (Catalão *et al.*, 2010). Nevertheless, it is not known if Gp1 only binds LysA inside the cell and then detaches from it and remains in the cytoplasm or if it is exported together with LysA. Therefore, to investigate the localization of Gp1 and LysA in *M. smegmatis* during an infection, we constructed a recombinant phage carrying Gp1 and LysA fused, at their C-terminus, to c-Myc and His_6 tags, respectively. This allowed us to detect and follow the production of both proteins by Western-blot in the course of an infection. A recombinant Ms6 phage carrying LysA fused to a His_6 tag at the C-terminus has already been generated for previous studies (Catalão *et al.*, 2010; Catalão *et al.*, 2011c). Thus, to construct the double mutant phage we only had to insert the c-Myc tag fused to the 3' end of *gp1*. For this purpose we employed the BRED technique recently developed by Marinelli *et al*. (2008). BRED system takes advantage of the proteins Gp60 and Gp61 from mycobacteriophage Che9c. Gp60 has an exonuclease activity that is dependent on the presence of dsDNA ends, whereas Gp61 is a ssDNA annealing protein recombinase that binds short (20 nucleotides) ssDNA as well as dsDNA substrates. The combined action of these proteins confers high levels of homologous recombination (van Kessel *et al.*, 2008). A targeting substrate was generated by amplifying a synthesized oligonucleotide containing the c-Myc tag sequence by PCR (see Material and Methods). The insertion was

accomplished in a single co-electroporation of Ms6 *lysA*-His_6 genomic DNA and a 210 bp targeting substrate containing *gp1* fused to the c-Myc sequence (Fig. 8a). In a total of 100 plaques screened, at least 4 contained the insertion, as identified by PCR using primers specific to the c-Myc sequence (Fig. 8b). Recombinant bacteriophages were plaque purified by several steps of serial dilutions in order to obtain pure mutant phages (Fig. 8c). DNA sequencing confirmed insertion of the c-Myc tag in frame with the C-terminus of Gp1 into Ms6 *lysA*-His_6 genome.



Recombinant Ms6 *gp1*-c-Myc *lysA*-His_6 phage

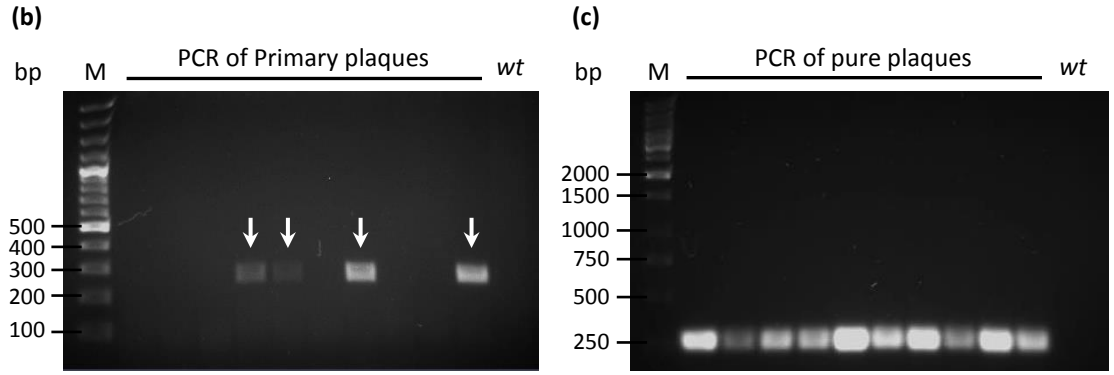


Figure 8. Construction of the recombinant Ms6 *gp1*-c-Myc *lysA*-His$_6$ phage. (a) Insertion of a c-Myc tag at the 3' end of *gp1* on Ms6 *lysA*-His$_6$ genome. A 110-mer oligonucleotide containing the c-Myc sequence was amplified by PCR to generate a 210-bp dsDNA substrate with homology regions to either side of the targeting sequence. Recombination between Ms6 *lysA*-His$_6$ phage genomic DNA and the 210-bp substrate generated the recombinant Ms6 *gp1*-c-Myc *lysA*-Hi$_6$ phage. (b) Gel electrophoresis of PCR products from primary plaques to detect phage Ms6 *gp1*-c-Myc *lysA*-His$_6$ with primers Prc-MyctagFwd and PrlysARv shows a band corresponding to the expected 252 bp. M – DNA Ladder: GeneRuler 100 bp Plus (Fermentas) (c) Mixed primary plaques were diluted and plated until only recombinant phages were detected by PCR. M – DNA Ladder: GeneRuler 1 kb (Fermentas); *wt* – wild-type Ms6 phage DNA; Arrows indicate primary plaques containing the desired recombinant phage.

The newly generated recombinant bacteriophage was shown to grow with normal plaque morphology and size (Fig. 9b). In addition, when comparing to the wild-type Ms6, one-step growth curves confirmed that the time of lysis and the number of phages released after infection remain unaffected by the presence of tag coding sequences at the end of Gp1 and LysA (Fig. 9a). This data indicated that we could use the recombinant Ms6 *gp1*-c-Myc *lysA*-His$_6$ to monitor Gp1 and LysA expression and localization during the course of an infection.

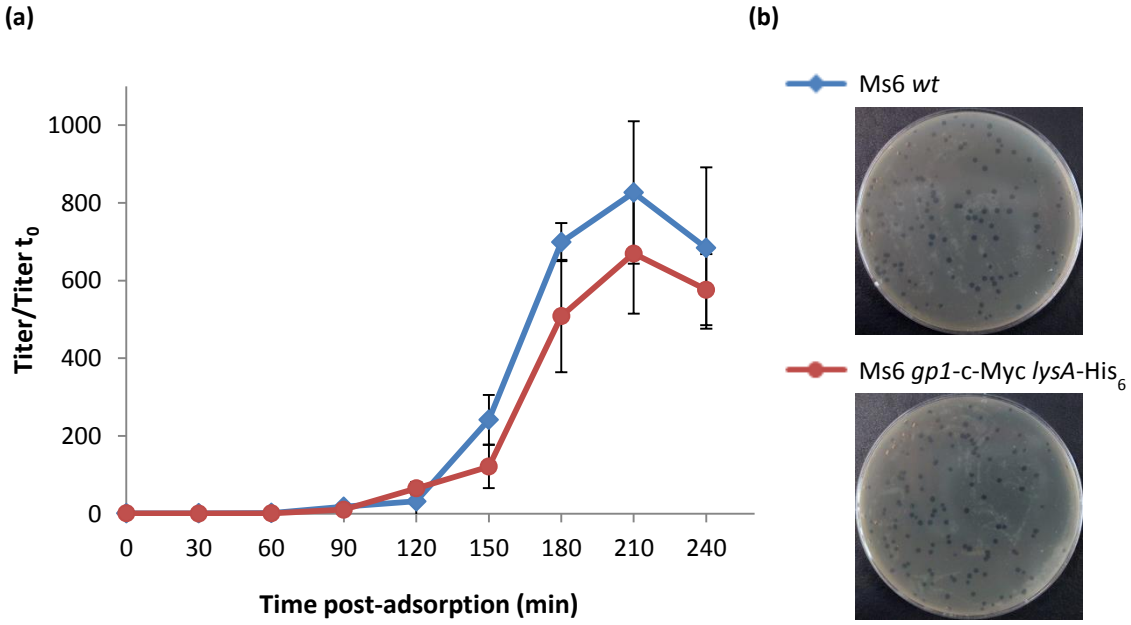


Figure 9. Evaluation of phage growth parameters. (a) One-step growth curves of phages Ms6 *wt* and Ms6 *gp1*-c-Myc *lysA*-His$_6$. For both phages, lysis starts to occur approximately 120 min after the 50-min phage adsorption period and there is no significant difference in the number of phage particles released after infection. Values plotted on the graph are an average of three independent experiments. Vertical bars represent the standard deviation. (b) Phage plaque morphology when infecting *M. smegmatis* cells with Ms6 *wt* and Ms6 *gp1*-c-Myc *lysA*-His$_6$.

2. Detection and localization of Gp1 and LysA in *M. smegmatis* infected cells

To detect the localization of Gp1 and LysA during the course of an infection, *M. smegmatis* was grown to mid-log phase and infected with mycobacteriophage Ms6 *gp1*-c-Myc *lysA*-His_6 at a MOI of 100. A high MOI was used to assure that the majority of cells were effectively infected, since our observations indicate that phage adsorption to mycobacteria cells is not very efficient. Samples were collected immediately before and every 30 min following adsorption for a period of 2 hours. Infected cells were lysed to generate whole-cell lysates, which were then subjected to differential ultracentrifugation to separate cell wall, membrane and cytosol-containing soluble fractions. The cell wall fraction consists of peptidoglycan covalently linked to arabinogalactan, which in turn is esterified by mycolic acids (Rezwan *et al.*, 2007), while the soluble fraction includes the cytosolic and the periplasmic content. The protein content of each subcellular fraction was estimated and the same amount of protein was separated by SDS-PAGE (Fig. 10a) and analysed by Western blot using antibodies that recognize the C-terminal c-Myc or His_6 epitope tags fused to Gp1 and LysA, respectively (Fig. 10b). Total protein staining reveals that each lane was loaded approximately with the same amount of protein. In addition, protein profiles are distinct between different fractions, which are in agreement with the subcellular fractionation protocol (Fig. 10a). As expected, using the anti-His_6 antibody, two proteins with 27.7 and 43.9 kDa were found, corresponding to the molecular weight of $Lysin_{241}$-His_6 and $Lysin_{384}$-His_6, respectively. Both proteins were first detected 90 min post-adsorption and were found solely on the cell wall fraction (Fig. 10b), which is not surprising since they target the peptidoglycan layer in order to compromise its integrity. Gp1 also started to be detectable 90 min following the 50 min adsorption period but in turn it is located on the cell wall and cell membrane fractions (Fig. 10b). MspA and KatG proteins were also detected to function as controls for the subcellular fractionation protocol. MspA is the main porin in *M. smegmatis* that enables the entry of small molecules through the permeability barrier

composed of the cell wall mycolates and other lipids. It assembles to an extremely stable oligomer with an apparent molecular weight of 100 kDa (Niederweis *et al.*, 1999). KatG is a cytosolic catalase-peroxidase from *M. tuberculosis* and its homologue in *M. smegmatis* has a molecular mass of 81.1 kDa (http://mycobrowser.epfl.ch/smegmalist.html). Due to their specific localization, these proteins are commonly used to monitor the effectiveness of the fractionation protocol (Rezwan *et al.*, 2007; Gibbons *et al.*, 2007). As expected, KatG is restricted to the soluble fraction, being undetectable on the cell envelope (Fig. 10b). In contrast, MspA is present on the wall fraction and it is also detectable, though in very small amounts, on the cytoplasmic membrane during the final time points of the infection (Fig. 10b). The presence of almost undetectable levels of MspA in the membrane fraction has also been reported by other authors (Rezwan *et al.*, 2007; Seeliger *et al.*, 2012). This validates the purity of the cell wall preparations and indicates that there was minimal contamination of membrane and soluble fractions with cell wall protein.

Overall, these results demonstrate that Gp1 is exported together with LysA to the extracytoplasmic environment, supporting the model that at lysis onset Gp1 behaves as a chaperone assisting LysA translocation to its target, the peptidoglycan cell wall. In addition localization of LysA is consistent with its peptidoglycan hydrolytic activity and with the presence of a PGRP domain.

3. Gp1 localization in *M. smegmatis* cells using translational fusions with PhoA'

Despite the analysis of Gp1 amino acid sequence did not predict the existence of a signal peptide; the results presented above have shown that Gp1 is exported, probably together with LysA, to the extracytoplasmic space, being present on the cell wall and cell membrane. Therefore, to address if Gp1 is endowed with a signal sequence we generated a hybrid protein where *gp1* is fused in frame with a signal sequence-less form of alkaline phosphatase (*phoA'*). Translational fusions to *E. coli phoA* have been widely used to study protein export

in bacteria because it is an exported enzyme that is enzymatically active only after its translocation into the bacterial periplasmic space. If a gene is fused to *E. coli phoA* lacking its own signal peptide sequence, only replacement sequences with protein export signals promote transfer across the cytoplasmic membrane, resulting in PhoA activity (Manoil and Beckwith, 1986). This approach has also been used to identify protein export sequences in *M. smegmatis* (Timm *et al.*, 1994).

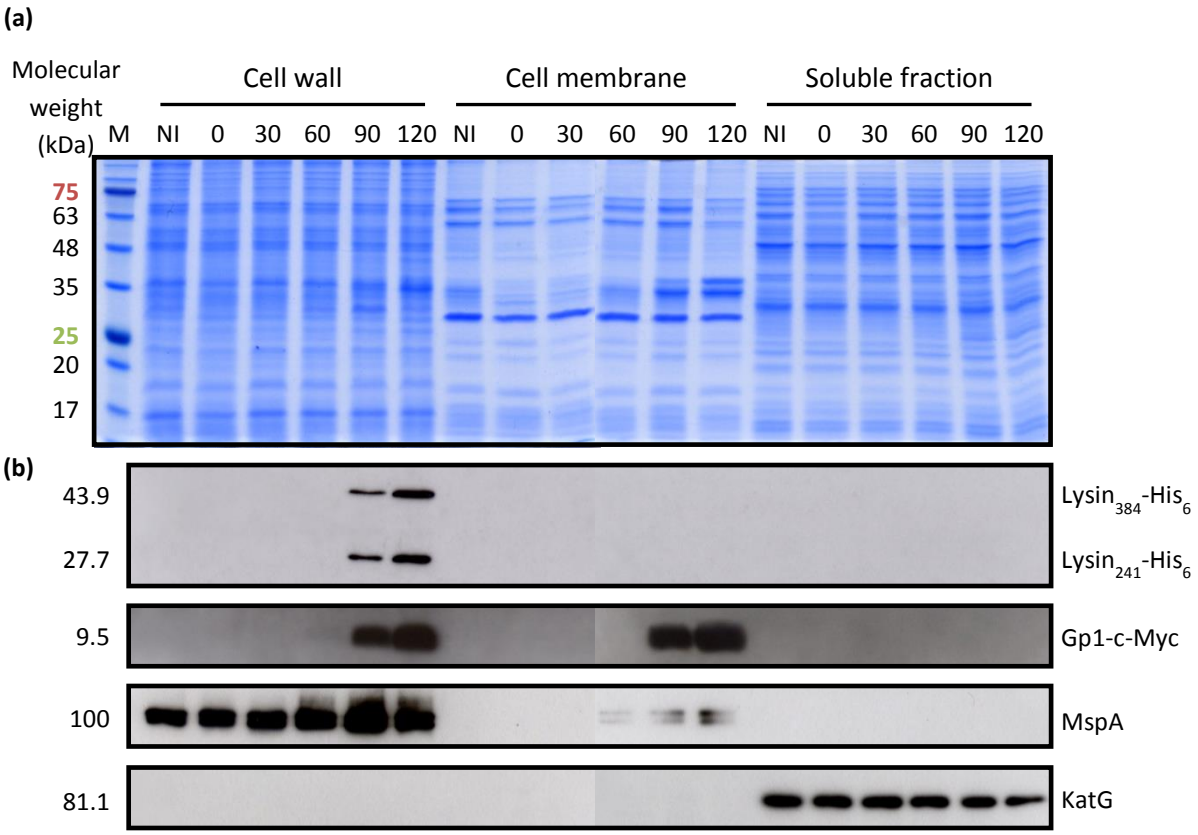


Figure 10. Subcellular fractionation of *M. smegmatis* infected cells. (a) *M. smegmatis* cells were infected with Ms6 *gp1*-c-Myc *lysA*-His_6 and samples were collected at different time points (numbers at the top of the gel indicate the time in minutes following the 50-min adsorption period). Samples were fractionated by differential ultracentrifugation and total protein was separated by SDS-PAGE. Total protein staining revealed that protein migration pattern is distinct between different cell compartments, which is consistent with the subcellular fractionation protocol. (b) Subcellular localization of Gp1-c-Myc and LysA-His_6 by immunoblot analysis. Gp1-c-Myc is detected both in the cell wall and cell membrane compartments, whereas LysA-His_6 is restricted to the cell wall fraction. The subcellular pattern of KatG and MspA expressed from the native chromosomal promoters was also evaluated. Each blot is representative of at least two independent experiments.

As Gp1 has growth inhibitory effects when it is overexpressed in *E. coli* (Catalão *et al.*, 2010), we started by trying to clone *gp1* fused to *phoA'* in pMP201 (M. Pimentel, unpublished data). pMP201 is an integrative mycobacterial vector developed in our lab in the past and transcription is driven by the promoter of the Ms6 lytic cassette. As a positive control we used the sequence encoding the first 50 aa of the *M. tuberculosis* 19-kDa antigen (*19-kDa$_{50aa}$*) fused to *phoA'*. The 19-kDa antigen is a glycosylated protein exported to the surface of *M. tuberculosis* (Herrmann *et al.*, 2000). It is synthesized with a Sec signal sequence, being processed by LspA, the lipoprotein signal peptidase (Sander *et al.*, 2004). Thus, by plating a strain carrying the 5' end of the 19-kDa antigen gene (*19-kDa$_{50aa}$*), which includes its signal sequence, fused to *phoA'* on plates containing the chromogenic phosphatase indicator BCIP, we would expect to obtain blue colonies. Neverthless, *M. smegmatis* cells carrying the *19kDa$_{50aa}$-phoA'* construct failed to yield blue colonies on these conditions (data not shown). This result was unexpected since 19-kDa antigen is known to be secreted and this kind of approach is widely used to study protein export in bacteria, including mycobacteria (Timm *et al.*, 1994; Lim *et al.*, 1995; Braunstein *et al.*, 2000). Actually, Herrmann *et al.* (1996) have previously shown, though using a different vector backbone, that this same construction has phosphatase activity on plates and our DNA sequencing analysis of the cloned fragments did not reveal any abnormality in the nucleotide sequence. We have not proceeded to investigate the reason why this construction did not work but the most likely explanation is related to protein expression problems.

To overcome this problem we decided to use another vector where expression of the hybrid proteins could be induced by the addition of acetamide to the culture medium. Thus we used pVVAP (Fig. 7), a replicative shuttle vector containing the inducible promoter of the acetamidase. pVVAP is an expression vector designed for mycobacteria, containing a RBS sequence as well as a GTG start codon downstream of the acetamidase promoter followed by a His_6 tag, which allows the synthesis of recombinant His_6 tagged proteins. We

constructed 3 plasmids derived from pVVAP carrying *gp1-phoA'*, *19-kDa*$_{50aa}$*-phoA'* or *phoA'* alone designated pFM1, pFM2 and pFM3, respectively (Fig. 11). When streaked on agar plates supplemented with BCIP, only *M. smegmatis* cells carrying plasmid pFM2 produced blue colonies, while cells harbouring pFM1 were colourless, suggesting that Gp1 does not contain an export signal sequence (Fig. 11).

To confirm this result PhoA activity was also measured in whole cells carrying the plasmids described above. Alkaline phosphatase activity can be assayed spectrophotometrically by quantifying *p*-nitrophenol, a yellow product that is generated when PhoA cleaves *p*-nitrophenyl phosphate (*p*NPP). Quantification of the alkaline phosphatase activity supported the results obtained on plates, revealing that cells harbouring pFM2 showed higher activity when compared to the other recombinants (Fig. 12a).

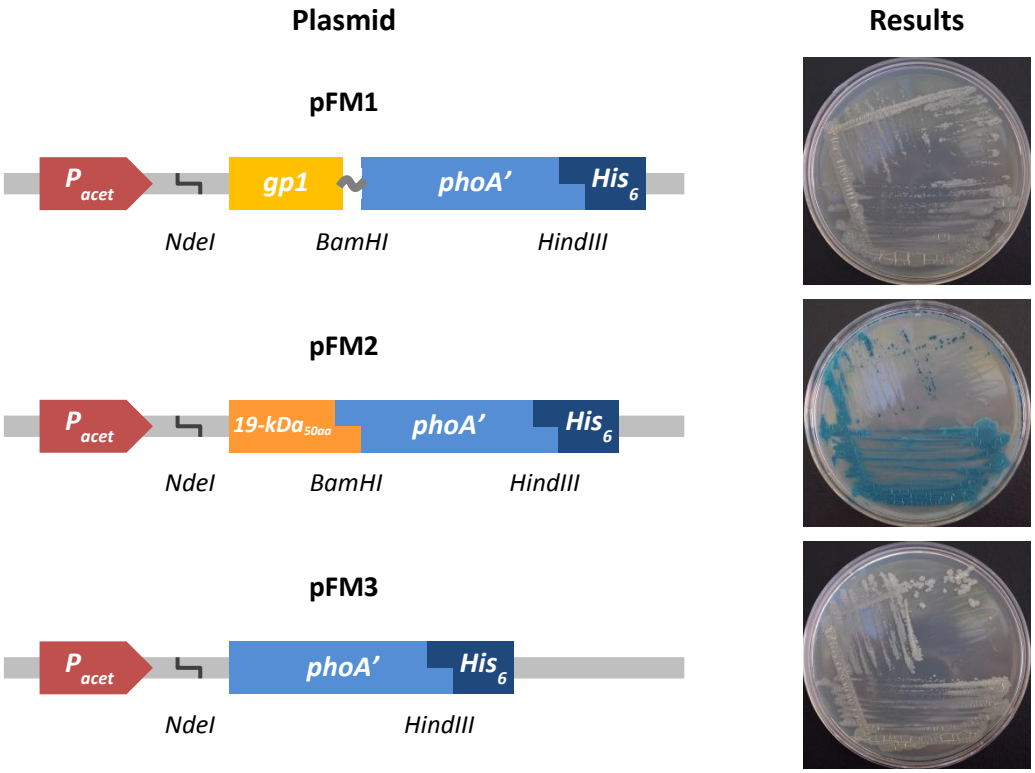


Figure 11. Alkaline phosphatase activity on BCIP-containing plates. Schematic representation of the plasmids used to transform *M. smegmatis* to evaluate the presence of an export signal on Gp1. Recombinant clones were plated on 7H9-agar supplemented with 15 µg mL^{-1} kanamycin, 0.2% Succinate, 0.2% Acetamide and 60 µg mL^{-1} BCIP to search for alkaline phosphatase activity. Pictures of plates after 7 days of incubation at 37ºC are shown on the right side of the panel. Plasmid pFM2 carrying the *19-kDa*$_{50aa}$ fused to *phoA'* was used as a positive control yielding blue colonies, as expected. P_{acet} – Acetamidase promoter.

Although unlikely, the lack of phosphatase activity observed could result from problems at the level of protein expression. To exclude this possibility we analysed the production of the recombinant proteins by Western-blot using an anti-His_6 antibody following SDS-PAGE. As seen in figure 12b, in inducing conditions PhoA' (pFM3) and the hybrid protein Gp1-PhoA' (pFM1) were produced with the expected molecular masses of 47.8 kDa and 56.5 kDa, respectively. Cells carrying $19kDa_{50aa}$-PhoA' (pFM2) generated two forms of the protein which would correspond to the full-length and mature proteins (resulting from signal peptide cleavage) with predicted molecular masses of 52.3 kDa and 50.3 kDa, respectively (Fig. 12b). Although the band that would correspond to the mature form of 19-kDa_{50aa}-PhoA' migrates with a lower molecular mass than what was expected, these results clearly demonstrate that the absence of phosphatase activity is not a result of an expression problem. Taken together, our data demonstrate that, despite being localized on the *M. smegmatis* cell wall and cell membrane, Gp1 is not endowed with a signal sequence that would allow its export.

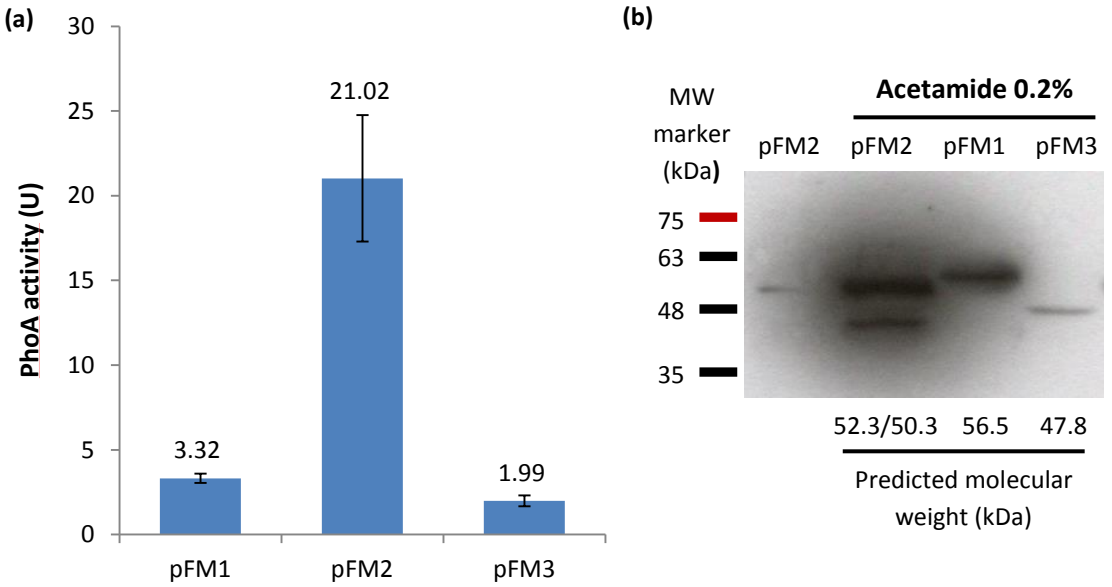
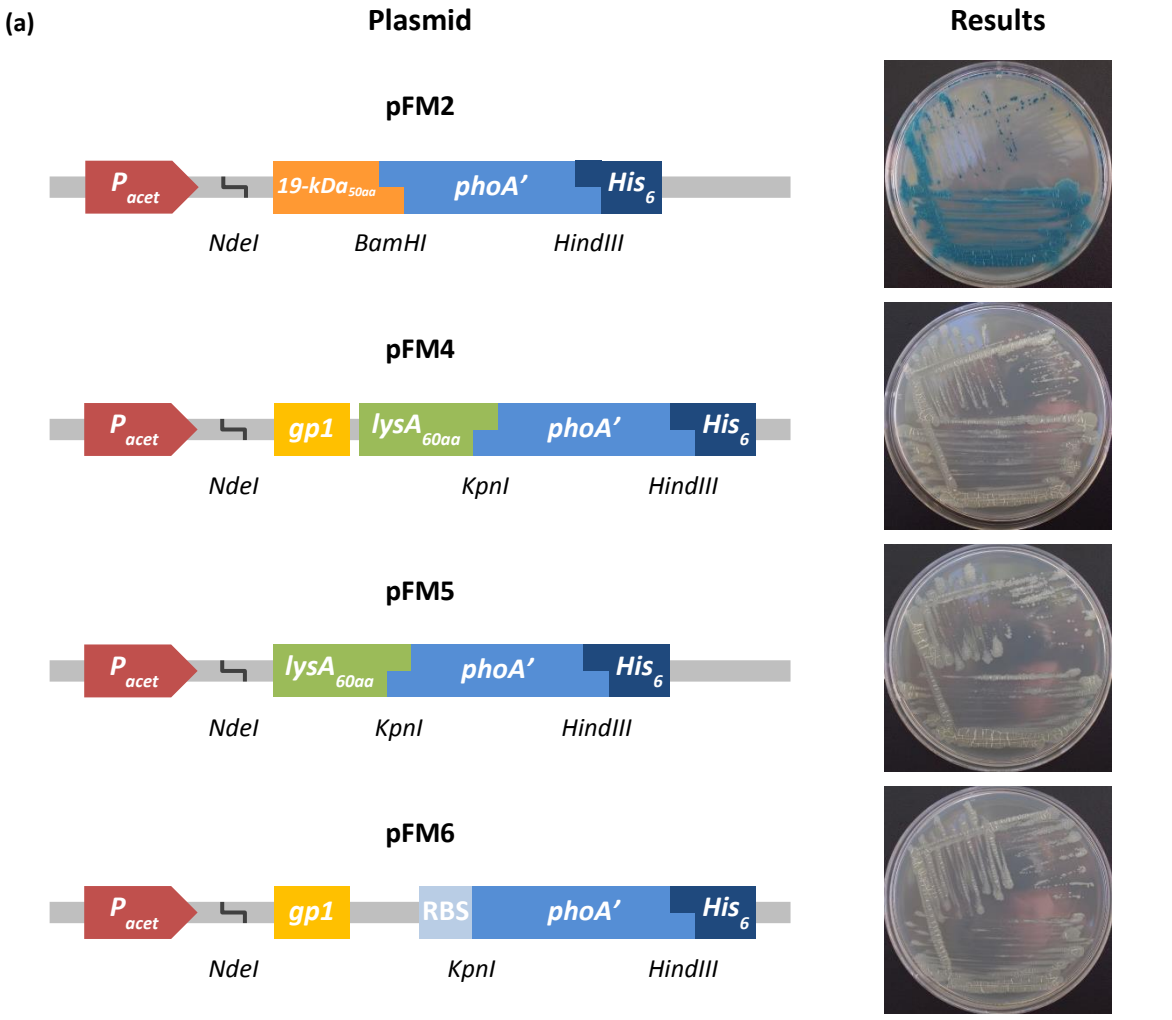


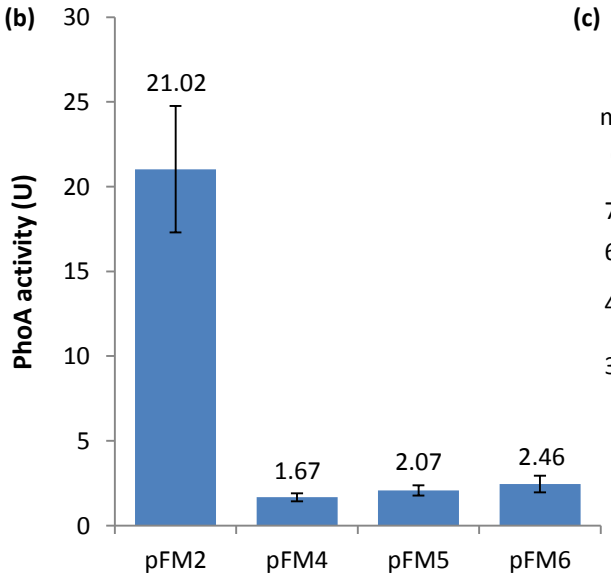
Figure 12. Detection of alkaline phosphatase. (a) The phosphatase activity in liquid cultures of *M. smegmatis* cells carrying pFM1, pFM2 and pFM3 was quantified using *p*NPP. Values are an average of two independent experiments. Vertical bars indicate the standard deviation. (b) Expression of the fusion proteins was confirmed by Western-blot. Cultures were grown to an OD_{600} of 0.4, induced with 0.2% acetamide for 6h and lysed by sonication (5 cycles of 15 s with 1 min resting on ice between cycles). A portion of the cell lysate was mixed with Laemmli buffer, boiled for 10 min, separated by SDS-PAGE, blotted onto a nitrocellulose membrane and probed with an anti-His_6 antibody.

4. Ability of Gp1 to translocate the first 60 aa of LysA to the extracytoplasmic environment

It has been shown that Gp1 is involved in endolysin transport across the cell inner membrane. LysA-PhoA' fusions were shown to be translocated to the extracytoplasmic environment only in the presence of Gp1. In addition, experiments in *E. coli* revealed that the first 60 aa of Ms6 LysA are necessary and sufficient for Gp1 binding, and are essential for LysA export (Catalão *et al.*, 2010). Since the N-terminal 60 aa of LysA are enough to mediate the interaction with Gp1, we hypothesized that this region could be enough to promote its export in *M. smegmatis*. To address this issue, we tested the ability of Gp1 to promote the translocation of PhoA' fused to the first 60 aa of LysA ($LysA_{60aa}$). For this purpose we constructed 4 plasmids derived from pVVAP carrying *gp1* and *$lysA_{60aa}$-phoA'* (pFM4), *$lysA_{60aa}$-phoA* (pFM5) or *phoA'* alone in the presence of *gp1* (pFM6) (Fig. 13a). To obey to the original gene disposition of Ms6, where the stop/start codons of Gp1 and LysA are overlapped we kept the same arrangement. Once more, we used plasmid pFM2 encoding the 19-kDa_{50aa} fused to PhoA' as a positive control. Recombinant *M. smegmatis* cells were plated on media containing BCIP to search for alkaline phosphatase activity. Only cells carrying pFM2 generated blue colonies, while the remaining recombinant clones were colourless (Fig. 13a).

Quantification of alkaline phosphatase activity in whole cells further confirmed the results obtained on plates and immunoblot analysis excluded protein expression issues, as in all constructs His_6-tagged proteins were detected by Western-blot (Fig. 13b and 13c). These results show the inability of Gp1 to promote the export of the LysA N-terminal 60 aa to the extracytoplasmic environment in *M. smegmatis*, suggesting the requirement for an additional factor for the process. Therefore, despite being essential, as previously demonstrated, LysA N-terminal 60 aa are not enough to mediate LysA export in *M. smegmatis*, in the presence of Gp1.


(a)
Plasmid
Results
pFM2
P acet
19-kDa 50aa
phoA'
His 6
NdeI
BamHI
HindIII
pFM4
P acet
gp1
lysA 60aa
phoA'
His 6
NdeI
KpnI
HindIII
pFM5
P acet
lysA 60aa
phoA'
His 6
NdeI
KpnI
HindIII
pFM6
P acet
gp1
RBS
phoA'
His 6
NdeI
KpnI
HindIII


(b)
PhoA activity (U)
30
25
20
15
10
5
0
21.02
1.67
2.07
2.46
pFM2
pFM4
pFM5
pFM6

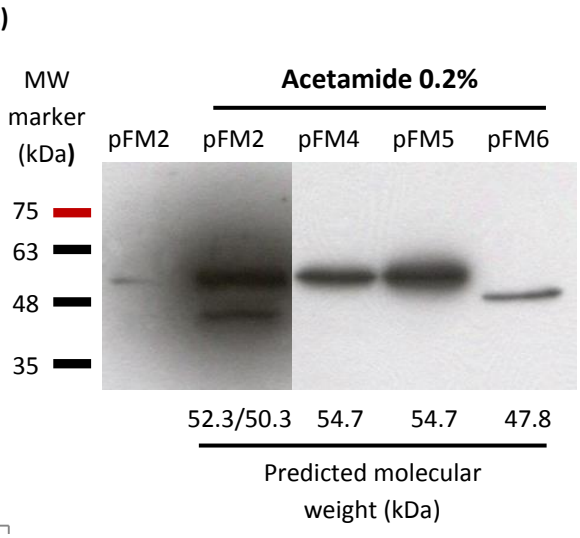

(c)
MW marker (kDa)
Acetamide 0.2%
pFM2
pFM2
pFM4
pFM5
pFM6
75
63
48
35
52.3/50.3
54.7
54.7
47.8
Predicted molecular weight (kDa)

Figure 13. Detection of alkaline phosphatase activity. (a) Schematic representation of the constructions used to transform *M. smegmatis* to evaluate the ability of Gp1 to translocate the N-terminal 60 aa of LysA across the cytoplasmic membrane. The colour of the colonies after 7 days of incubation at 37ºC on 7H9-agar containing 15 μg mL^{-1} kanamycin, 0.2% Succinate, 0.2% Acetamide and 60 μg mL^{-1} BCIP is shown on the right side of the figure. pFM2 was used as a positive control. (b) The phosphatase activity in liquid cultures of *M. smegmatis* cells carrying pFM2, pFM4, pFM5 and pFM6 was determined using *p*NPP. Values are an average of two independent experiments. Vertical bars indicate the standard deviation. (c) Expression of the fusion proteins was confirmed by Western-blot. Cultures were grown to an OD_{600} of 0.4, induced with 0.2% acetamide for 6h and lysed as described above. A portion of the cell lysate was mixed with Laemmli buffer, boiled for 10 min, separated by SDS-PAGE, blotted onto a nitrocellulose membrane and probed with an anti-His_6 antibody. P_{acet} – acetamidase promoter.

IV. DISCUSSION

In this work we have explored the localization of two phage proteins involved in Ms6 mediated lysis, Gp1 and LysA. Our group has recently described that Gp1 is a chaperone-like protein that binds the N-terminal region of $LysA_{384}$, assisting its export across the cytoplasmic membrane in a holin-independent manner (Catalão *et al.*, 2010). Gp1 is encoded by the first gene in the Ms6 lytic operon, *gp1*, which is located immediately upstream of *lysA* and overlaps *lysA* start codon in a different reading frame. In turn, *lysA* encodes two products in the same reading frame with endolysin activity, $Lysin_{384}$ and $Lysin_{241}$. $Lysin_{241}$ results from a second translation event within *lysA* and consequently the N-terminal region that interacts with Gp1 is absent (Catalão *et al.*, 2011c).

Here we provide evidence that Gp1 and both endolysins, $LysA_{384}$ and $LysA_{241}$, are associated with the cell envelope of *M. smegmatis* infected cells. Despite the fact that available bioinformatic programs did not predict Gp1 to be endowed with a signal sequence, subcellular fractionation assays have shown that this chaperone-like protein is located on the cell wall and cell membrane. This led us to construct a Gp1-PhoA' hybrid protein to investigate the presence of a putative signal peptide. The absence of alkaline phosphatase activity in cells expressing the fusion Gp1-PhoA' shows that Gp1 is not exported across the cytoplasmic membrane, indicating that it is not endowed with a signal sequence. Export of mycobacterial proteins without recognizable signal exports have already been described (Rigel and Braunstein, 2008) and one must keep in mind that mycobacteria export mechanisms are not fully understood. With these observations in mind we raised the

question if Gp1 and LysA must be associated to reach their final destination in order to accomplish their biological function. This strategy resembles characteristics of mycobacteria's export mechanisms and type III and IV secretion systems as many secreted virulence determinants of pathogenic bacteria by these pathways are bacteriophage-encoded proteins (Miao and Miller, 1999).

Mycobacteria employ different pathways to export bacterial products across their extremely hydrophobic and thick cell envelope. The general secretion pathway and the Twin-Arginine Transporter (TAT) are examples of these mechanisms; however these systems require the existence of a signal sequence in the proteins targeted for export. More recently, a specialized protein secretion pathway termed ESX-1 or *Snm* (secretion in mycobacteria) was identified in *M. tuberculosis* (Champion and Cox, 2007). ESAT-6 (Early secreted antigen target, 6 kDa) and CFP-10 (Culture filtrate protein, 10 kDa) are two potent immune response elicitors secreted through this pathway. These two proteins lack obvious signal sequences and interact with each other, just like Gp1 and LysA, to form a tight dimer (Renshaw *et al.*, 2002; Stanley *et al.*, 2003). Secretion of ESAT-6 depends on the presence of CFP-10, and in the mycobacterial cell, they are interdependent on each other for stability (Champion and Cox, 2007). Indeed, previous experiments performed in our lab have also shown that synthesis and/or stability of the larger endolysin (Lys_{384}) is highly dependent on Gp1 production (Catalão *et al.*, 2011c). A reasonable explanation could be that the endolysin becomes unstable in the cytoplasm in the absence of its chaperone. It seems that there are significant parallels at the molecular level between the way Gp1-LysA complex reaches the extracytoplasmic environment, the ESX-1 system and the Type III and IV secretion pathways (Champion and Cox, 2007). Our data suggest that Gp1 and LysA might be targeted for export as a substrate pair, similarly to what happens with ESAT-6 and CFP-10. Indeed, our subcellular fractionation results show that Gp1 and LysA are both targeted to the cell envelope at the same time. In addition, previous results have shown that a LysA-PhoA'

fusion is exported to the periplasm only in the presence of Gp1 (Catalão *et al.*, 2010), which reinforces the idea that Gp1 and LysA are dependent on each other. In fact Type IV secretion systems have examples of chaperone-substrate pairs that are targeted and secreted from the bacterial cell (Sundberg and Ream, 1999; Duménil and Isberg, 2001). Specifically, like Gp1, the chaperones of these systems are small proteins (10-15 kDa) with an acidic pI and act as dimers, binding to the N-terminal region of the cognate protein. In addition, it is almost a rule that the chaperones are encoded adjacent to the effector. These chaperones typically function to keep the substrate in a secretion competent conformation and to prevent interaction with other proteins or aggregation (Champion and Cox, 2007). In the particular case of mycobacteriophage Ms6 we believe that besides assisting LysA export, Gp1 may have an additional role in the regulation of LysA activity.

It was previously determined that the N-terminal 60 aa of Ms6 LysA are necessary and sufficient for Gp1 binding and are essential for LysA export (Catalão *et al.*, 2010). In this study we have also tested if this 60 aa sequence was enough to promote the export of *E. coli* PhoA lacking its signal sequence. Our results have shown that, despite being essential for LysA export, LysA N-terminal sequence is not sufficient to promote its own translocation across the cytoplasmic membrane in the presence of Gp1. As mentioned previously, in the presence of Gp1 the full-length LysA fused to PhoA' is exported to the extracytoplasmic environment. Hence, there must be another factor, in addition to the region encompassing Gp1 binding site, crucial for LysA export. This observation together with the previous idea that Gp1 and LysA may be targeted for export as a substrate pair imposes a central role in the formation of the Gp1-LysA complex, rather than in specific sequences on either of the proteins. In fact, it has already been described that TTS chaperones bound to the effector protein could function as 3D export signals (Cornelis, 2006). One of those examples is the targeting of the TTS EspA filament protein and its cognate chaperone, CesAB, to the injectisome of the enteropathogenic *E. coli*. In that study the authors have shown that the

EscN ATPase, a key protein in TTS systems located at the entrance of the injectisome, selectively engages the EspA-loaded CesAB but not the unliganded CesAB. Structural analysis revealed that the targeting signal is encoded in a conformational switch in the chaperone that is induced only upon binding to the physiological substrate (Chen *et al.*, 2013). Although a similar mechanism could be employed during Ms6 mediated lysis, this idea is purely speculative and additional studies are required to elucidate this aspect.

Although one step-growth curves reveal that lysis starts to occur approximately 120 minutes after phage adsorption, our data from subcellular fractionation assays indicate that Gp1 and LysA start to be detectable 90 minutes post-adsorption and are absent from the soluble fraction. In fact, we have found that LysA is restricted to the cell wall, which is not surprising since it targets the murein layer and holds a central PGRP conserved domain. This domain lies between amino acid residues 168 and 312, which explains the association of both $LysA_{241}$ and $LysA_{384}$ with the cell wall (Catalão *et al.*, 2011c). PGRPs (cd06583) are pattern recognition receptors that bind, and in certain cases, hydrolyze the peptidoglycan of bacterial cell walls (Dziarski, 2004). These data together with previous observations of complete lysis of *M. smegmatis* cells expressing Gp1 and LysA after a treatment with nisin suggest that both proteins are not accumulated in the cytoplasm of the cell as it happens with lambda phage (Catalão *et al.*, 2010). Instead, Gp1 and LysA seem to be positioned at their final destination as they are being synthesized rather than at the end of phage maturation. This strategy might be particularly advantageous in the case of mycobacteria due to the complex structure of their cell walls. Evolutionary pressure should favor an optimum balance between the duration of a lytic cycle and effective progeny yield. If lysis is premature, no or very few phages will have been produced; but if delayed it compromises the opportunity to infect new hosts. Thus, building up a lytic arsenal at their site of action as phage assembly takes place may be a sensible strategy for phages infecting hosts with thick murein walls in order to assure a quick cell lysis once an adequate number of progeny virions

is reached intracellularly (São-José *et al.*, 2000). Though, in the case of secreted endolysins some additional extracytoplasmic regulatory mechanism must be operative to ensure that premature lysis does not occur. In the case of *E. coli* phages P1 and 21, endolysins possess an N-terminal SAR sequence that allows the enzyme to be exported to the membrane where it is arrested (Xu *et al.*, 2004; Park *et al.*, 2007). When lysis is triggered by the cognate holins, through dissipation of the pmf, lysins are released as soluble active enzymes in the periplasm and access the peptidoglycan to accomplish host-cell lysis (Xu *et al.*, 2005; Sun *et al.*, 2009). Similarly, in Ms6 there must be some kind of mechanism that restrains LysA activity until the proper time of lysis. We speculate that Gp1 might be involved in the maintenance of $Lysin_{384}$ in an inactive state until lysis is triggered. The association observed between Gp1 and the cell envelope, especially with the cell membrane, supports this idea. Gp1 binding to the N-terminal domain may alter endolysin conformation and block substrate binding or may prevent the access of LysA to the peptidoglycan substrate. Accordingly, it is hypothesized that Ms6 holin has merely a depolarizing role, as described for the lambdoid phage 21 pinholin ($S^{21}68$). When $S^{21}68$ triggers, it eliminates the pmf, causing endolysin activation, but does not form holes in the membrane, large enough to allow passage of the cytoplasmic endolysin (Park *et al.*, 2007). The fact that LysA is exported together with Gp1 strengthens the idea that LysA activation depends on the energized state of the cytoplasmic membrane, as described for secreted endolysins. However, further studies are required to depict the detailed molecular mechanisms underlying LysA export and subsequent activation.

A clear relationship between Gp1 and LysA has been demonstrated in Ms6, however the same does not seem to occur among all other mycobacteriophages sequenced so far. Gp1 homologues have been identified in other mycobacteriophage genomes, particularly in the lysis cassette of phages that encode two endolysins. Interestingly, however, two putative translational initiation signals were also identified in endolysin genes belonging to five

mycobacteriophages (Phyler, Phaedrus, Pipefish, Corndog and LeBron) that do not possess Gp1 homologues but possess N-terminal related Ms6 LysA sequences. The lack of representation of Gp1 homologues upstream of *lysA* in these phages could result from loss of *gp1*-like genes in these genomes and suggests that endolysin export occurs in a different way. In addition, in three mycobacteriophages (TM4, Jasper and Lockley) that possess Gp1 similar proteins but unrelated Ms6 LysA enzyme, this *lysA* gene arrangement was not observed, suggesting that *gp1*-like genes in these mycobacteriophages might result from recent acquisition by horizontal genetic exchange (Hatfull *et al.*, 2010). Gp1 may confer a selective advantage for host cell lysis under different environmental conditions: very small differences in lysis timing and efficiency are strongly selective because of competition for hosts by new released progeny (Young, 2005). Although Ms6 produces two endolysins and a chaperone-like protein, whose functions were shown to be important for a complete and efficient lysis, some questions remain to be answered: Why do some mycobacteriophages need to produce two endolysins and Gp1 homologues? Is this phenomenon only a consequence of gene transfer throughout evolution? More studies with other mycobacteriophages will certainly help to clarify the need for these genes.

V. REFERENCES

Ackermann, H. W. (2003) Bacteriophage observations and evolution. *Res Microbiol* **154**: 245–51.

Ackermann, H. W. (2009) Phage classification and characterization. *Methods Mol Biol* **501**: 127–40.

Bernhardt, T.G., Struck, D.K., and Young, R. (2001) The lysis protein E of phi X174 is a specific inhibitor of the MraY-catalyzed step in peptidoglycan synthesis. *J Biol Chem* **276**: 6093–7.

Bernhardt, T.G., Wang, I. N., Struck, D.K., and Young, R. (2002) Breaking free: "protein antibiotics" and phage lysis. *Res Microbiol* **153**: 493–501.

Berry, J., Rajaure, M., Pang, T., and Young, R. (2012) The spanin complex is essential for lambda lysis. *J Bacteriol* **194**: 5667–74.

Berry, J., Savva, C., Holzenburg, A., and Young, R. (2010) The lambda spanin components Rz and Rz1 undergo tertiary and quaternary rearrangements upon complex formation. *Protein Sci* **19**: 1967–77.

Berry, J., Summer, E.J., Struck, D.K., and Young, R. (2008) The final step in the phage infection cycle: the Rz and Rz1 lysis proteins link the inner and outer membranes. *Mol Microbiol* **70**: 341–51.

Bläsi, U., and Young, R. (1996) Two beginnings for a single purpose: the dual-start holins in the regulation of phage lysis. *Mol Microbiol* **21**: 675–82.

Braunstein, M., Griffin TJ, I. V, Kriakov, J.I., Friedman, S.T., Grindley, N.D., and Jacobs, W.R. (2000) Identification of genes encoding exported *Mycobacterium tuberculosis* proteins using a Tn552'phoA in vitro transposition system. *J Bacteriol* **182**: 2732–40.

Brennan, P.J. (2003) Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **83**: 91–7.

Briers, Y., Volckaert, G., Cornelissen, A., Lagaert, S., Michiels, C.W., Hertveldt, K., and Lavigne, R. (2007) Muralytic activity and modular structure of the endolysins of *Pseudomonas aeruginosa* bacteriophages phiKZ and EL. *Mol Microbiol* **65**: 1334–44.

Brüssow, H., Canchaya, C., and Hardt, W. D. (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* **68**: 560–602.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L., and Brüssow, H. (2003) Phage as agents of lateral gene transfer. *Curr Opin Microbiol* **6**: 417–24.

Catalão, M.J., Gil, F., Moniz-Pereira, J., and Pimentel, M. (2010) The mycobacteriophage Ms6 encodes a chaperone-like protein involved in the endolysin delivery to the peptidoglycan. *Mol Microbiol* **77**: 672–86.

Catalão, M.J., Gil, F., Moniz-Pereira, J., and Pimentel, M. (2011a) Functional analysis of the holin-like proteins of mycobacteriophage Ms6. *J Bacteriol* **193**: 2793–803.

Catalão, M.J., Gil, F., Moniz-Pereira, J., and Pimentel, M. (2011b) The endolysin-binding domain encompasses the N-terminal region of the mycobacteriophage Ms6 Gp1 chaperone. *J Bacteriol* **193**: 5002–6.

Catalão, M. J., Gil, F., Moniz-Pereira, J., São-José, C., and Pimentel, M. (2013) Diversity in bacterial lysis systems: bacteriophages show the way. *FEMS Microbiol Rev* **37**: 554–71.

Catalão, M.J., Milho, C., Gil, F., Moniz-Pereira, J., and Pimentel, M. (2011c) A second endolysin gene is fully embedded in-frame with the lysA gene of mycobacteriophage Ms6. *PLoS One* **6**: e20515.

Champion, P.A.D., and Cox, J.S. (2007) Protein secretion systems in Mycobacteria. *Cell Microbiol* **9**: 1376–84.

Chen, L., Ai, X., Portaliou, A.G., Minetti, C. a S. a, Remeta, D.P., Economou, A., and Kalodimos, C.G. (2013) Substrate-activated conformational switch on chaperones encodes a targeting signal in type III secretion. *Cell Rep* **3**: 709–15.

Cornelis, G.R. (2006) The type III secretion injectisome. *Nat Rev Microbiol* **4**: 811–25.

Daffé, M., Dupon, M., and Gas, N. (1989) The cell envelope of *Mycobacterium smegmatis*: cytochemistry and architectural implications. *FEMS Microbiol Lett* **52**: 89–93.

Dewey, J.S., Savva, C.G., White, R.L., Vitha, S., Holzenburg, A., and Young, R. (2010) Micron-scale holes terminate the phage infection cycle. *Proc Natl Acad Sci U S A* **107**: 2219–23.

Duménil, G., and Isberg, R.R. (2001) The *Legionella pneumophila* IcmR protein exhibits chaperone activity for IcmQ by preventing its participation in high-molecular-weight complexes. *Mol Microbiol* **40**: 1113–27.

Dziarski, R. (2004) Peptidoglycan recognition proteins (PGRPs). *Mol Immunol* **40**: 877–86.

Freitas-Vieira, A., Anes, E., and Moniz-Pereira, J. (1998) The site-specific recombination locus of mycobacteriophage Ms6 determines DNA integration at the tRNA(Ala) gene of *Mycobacterium* spp. *Microbiology* **144**: 3397–406.

Garcia, M., Pimentel, M., and Moniz-Pereira, J. (2002) Expression of Mycobacteriophage Ms6 lysis genes is driven by two sigma(70)-like promoters and is dependent on a transcription termination signal present in the leader RNA. *J Bacteriol* **184**: 3034–43.

Gibbons, H.S., Wolschendorf, F., Abshire, M., Niederweis, M., and Braunstein, M. (2007) Identification of two *Mycobacterium smegmatis* lipoproteins exported by a SecA2-dependent pathway. *J Bacteriol* **189**: 5090–100.

Gil, F., Catalão, M.J., Moniz-Pereira, J., Leandro, P., McNeil, M., and Pimentel, M. (2008) The lytic cassette of mycobacteriophage Ms6 encodes an enzyme with lipolytic activity. *Microbiology* **154**: 1364–71.

Gil, F., Grzegorzewicz, A.E., Catalão, M.J., Vital, J., McNeil, M.R., and Pimentel, M. (2010) Mycobacteriophage Ms6 LysB specifically targets the outer membrane of *Mycobacterium smegmatis*. *Microbiology* **156**: 1497–504.

Gründling, A., Manson, M.D., and Young, R. (2001) Holins kill without warning. *Proc Natl Acad Sci U S A* **98**: 9348–52.

Hatfull, G.F. (2000) Molecular genetics of mycobacteriophages. In *Molecular genetics of mycobacteria*. Hatfull, G.F., and Jacobs, W.R. (eds). American Society for Microbiology, Washington, DC. pp. 3–54.

Hatfull, G.F. (2006) Mycobacteriophages. In *The Bacteriophages*. Calendar, R. (ed.). Oxford University Press, New York. pp. 602–620.

Hatfull, G.F. (2010) Mycobacteriophages: genes and genomes. *Annu Rev Microbiol* **64**: 331–56.

Hatfull, G.F. (2012a) *The secret lives of mycobacteriophages. Adv Virus Res* **82**: 179-288.

Hatfull, G.F. (2012b) Complete genome sequences of 138 mycobacteriophages. *J Virol* **86**: 2382–4.

Hatfull, G.F., and Jacobs, W.R. (1994) Mycobacteriophages: cornerstones of mycobacterial research. In *Tuberculosis: pathogenesis, protection and control*. Bloom, B. (ed.). Amer Society for Microbiology, Washington DC. pp. 165–183.

Hatfull, G.F., Jacobs-Sera, D., Lawrence, J.G., Pope, W.H., Russell, D.A., Ko, C.-C., *et al.* (2010) Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *J Mol Biol* **397**: 119–43.

Hendrix, R.W. (2003) Bacteriophage genomics. *Curr Opin Microbiol* **6**: 506–11.

Herrmann, J.L., Delahay, R., Gallagher, A., Robertson, B., and Young, D. (2000) Analysis of post-translational modification of mycobacterial proteins using a cassette expression system. *FEBS Lett* **473**: 358–62.

Herrmann, J.L., O'Gaora, P., Gallagher, a, Thole, J.E., and Young, D.B. (1996) Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from *Mycobacterium tuberculosis*. *EMBO J* **15**: 3547–54.

Hoffmann, C., Leis, A., Niederweis, M., Plitzko, J.M., and Engelhardt, H. (2008) Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proc Natl Acad Sci U S A* **105**: 3963–7.

Kutateladze, M., and Adamia, R. (2010) Bacteriophages as potential new therapeutics to replace or supplement antibiotics. *Trends Biotechnol* **28**: 591–5.

Kuty, G.F., Xu, M., Struck, D.K., Summer, E.J., and Young, R. (2010) Regulation of a phage endolysin by disulfide caging. *J Bacteriol* **192**: 5682–7.

Lemassu, A., and Daffé, M. (1994) Structural features of the exocellular polysaccharides of *Mycobacterium tuberculosis. Biochem J* **297**: 351–7.

Lemassu, A., Ortalo-Magné, A., Bardou, F., Silve, G., Laneélle, M.A., and Daffé, M. (1996) Extracellular and surface-exposed polysaccharides of non-tuberculous mycobacteria. *Microbiology* **142**: 1513–20.

Lim, E.M., Rauzier, J., Timm, J., Torrea, G., Murray, A., Gicquel, B., and Portnoi, D. (1995) Identification of *Mycobacterium tuberculosis* DNA sequences encoding exported proteins by using phoA gene fusions. *J Bacteriol* **177**: 59–65.

Loessner, M.J. (2005) Bacteriophage endolysins--current state of research and applications. *Curr Opin Microbiol* **8**: 480–7.

Loessner, M.J., Kramer, K., Ebel, F., and Scherer, S. (2002) C-terminal domains of *Listeria monocytogenes* bacteriophage murein hydrolases determine specific recognition and high-affinity binding to bacterial cell wall carbohydrates. *Mol Microbiol* **44**: 335–49.

Madigan, M.T., Martinko, J.M., Sthal, D., and Clark, D. (2010) *Brock Biology of Microorganisms*. 13th ed., Benjamin Cummings, San Francisco.

Manoil, C., and Beckwith, J. (1986) A genetic approach to analyzing membrane protein topology. *Science* **233**: 1403–8.

Marinelli, L.J., Piuri, M., Swigonová, Z., Balachandran, A., Oldfield, L.M., Kessel, J.C. van, and Hatfull, G.F. (2008) BRED: a simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. *PLoS One* **3**: e3957.

Miao, E.A., and Miller, S.I. (1999) Bacteriophages in the evolution of pathogen-host interactions. *Proc Natl Acad Sci U S A* **96**: 9452–4.

Niederweis, M., Ehrt, S., Heinz, C., Klöcker, U., Karosi, S., Swiderek, K.M., *et al.* (1999) Cloning of the mspA gene encoding a porin from *Mycobacterium smegmatis*. *Mol Microbiol* **33**: 933–45.

Parish, T., Mahenthiralingam, E., Draper, P., Davis, E.O., and Colston, M.J. (1997) Regulation of the inducible acetamidase gene of *Mycobacterium smegmatis. Microbiology **143:** 2267–76*

Pimentel, M. (1999) Biologia molecular da resistência à superinfecção por bacteriófagos: estudo do gene pin do micobacteriófago Ms6. PhD Thesis. University of Lisbon

Portugal, I., Anes, E., and Moniz-Pereira, J. (1989) Temperate mycobacteriophage from *M. smegmatis*. *Acta Leprol* **7 Suppl 1**: 243–4.

Renshaw, P.S., Panagiotidou, P., Whelan, A., Gordon, S. V, Hewinson, R.G., Williamson, R.A., and Carr, M.D. (2002) Conclusive evidence that the major T-cell antigens of the *Mycobacterium tuberculosis* complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6*CFP-10 complex. Implications for pathogenesis and virulence. *J Biol Chem* **277**: 21598–603.

Rezwan, M., Lanéelle, M.-A., Sander, P., and Daffé, M. (2007) Breaking down the wall: fractionation of mycobacteria. *J Microbiol Methods* **68**: 32–9.

Rigel, N.W., and Braunstein, M. (2008) A new twist on an old pathway--accessory Sec [corrected] systems. *Mol Microbiol* **69**: 291–302.

Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Rodriguez-Brito, B., Pasić, L., Thingstad, T.F., Rohwer, F., and Mira, A. (2009) Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* **7**: 828–36.

Russel, M. (1995) Moving through the membrane with filamentous phages. *Trends Microbiol* **3**: 223–8.

Sambrook, J., and Russel, D. (2001) *Molecular Cloning: A Laboratory Manual*. 3rd ed., Cold Spring Harbor Press, New York.

Sander, P., Rezwan, M., Walker, B., Rampini, S.K., Kroppenstedt, R.M., Ehlers, S., *et al.* (2004) Lipoprotein processing is required for virulence of *Mycobacterium tuberculosis. Mol Microbiol* **52**: 1543–52.

Sani, M., Houben, E.N.G., Geurtsen, J., Pierson, J., Punder, K. de, Zon, M. van, *et al.* (2010) Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile ,structure containing ESX-1-secreted proteins. *PLoS Pathog* **6**: e1000794.

São-José, C., Nascimento, J., Parreira, R., and Santos, M. (2007) Release of progeny phages from infected cells. In *Bacteriophage: Genetics and Molecular Biology*. McGrath, S., and Sinderen, D. van (eds). Caister Academic Press, Norfolk. pp. 307–333.

São-José, C., Parreira, R., Vieira, G., and Santos, M.A. (2000) The N-terminal region of the Oenococcus oeni bacteriophage fOg44 lysin behaves as a bona fide signal peptide in Escherichia coli and as a cis-inhibitory element, preventing lytic activity on oenococcal cells. *J Bacteriol* **182**: 5823–31.

Schmelcher, M., Donovan, D.M., and Loessner, M.J. (2012) Bacteriophage endolysins as novel antimicrobials. *Future Microbiol* **7**: 1147–71.

Seeliger, J.C., Holsclaw, C.M., Schelle, M.W., Botyanszki, Z., Gilmore, S.A., Tully, S.E., *et al.* (2012) Elucidation and chemical modulation of sulfolipid-1 biosynthesis in *Mycobacterium tuberculosis*. *J Biol Chem* **287**: 7990–8000.

Smith, M., Jessee, J., Landers, T., and Jordan, J. (1990) High efficiency bacterial electroporation: 1 x 1010 *E. coli* transformants/µg. *Focus (Madison)* **12**.

Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T., and Jacobs, W.R. (1990) Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* **4**: 1911–9.

Snyder, L., and Champness, W. (2007) *Molecular Genetics of Bacteria.* 3rd ed., American Society for Microbiology, Washington, DC.

Stanley, S.A., Raghavan, S., Hwang, W.W., and Cox, J.S. (2003) Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc Natl Acad Sci U S A* **100**: 13001–6.

Sun, Q., Kuty, G.F., Arockiasamy, A., Xu, M., Young, R., and Sacchettini, J.C. (2009) Regulation of a muralytic enzyme by dynamic membrane topology. *Nat Struct Mol Biol* **16**: 1192–4.

Sundberg, C.D., and Ream, W. (1999) The *Agrobacterium tumefaciens* chaperone-like protein, VirE1, interacts with VirE2 at domains required for single-stranded DNA binding and cooperative interaction. *J Bacteriol* **181**: 6850–5.

Timm, J., Perilli, M.G., Duez, C., Trias, J., Orefici, G., Fattorini, L., *et al.* (1994) Transcription and expression analysis, using lacZ and phoA gene fusions, of *Mycobacterium fortuitum* beta-lactamase genes cloned from a natural isolate and a high-level beta-lactamase producer. *Mol Microbiol* **12**: 491–504.

van Kessel, J.C., and Hatfull, G.F. (2007) Recombineering in *Mycobacterium tuberculosis*. *Nat Methods* **4**: 147–52.

van Kessel, J.C., Marinelli, L.J., and Hatfull, G.F. (2008) Recombineering mycobacteria and their phages. *Nat Rev Microbiol* **6**: 851–7.

Walmagh, M., Briers, Y., Santos, S.B. dos, Azeredo, J., and Lavigne, R. (2012) Characterization of modular bacteriophage endolysins from Myoviridae phages OBP, 201ϕ2-1 and PVP-SE1. *PLoS One* **7**: e36991.

Weinbauer, M.G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28**: 127–81.

White, R., Chiba, S., Pang, T., Dewey, J.S., Savva, C.G., Holzenburg, A., *et al.* (2011) Holin triggering in real time. *Proc Natl Acad Sci U S A* **108**: 798–803.

Wolschendorf, F., Mahfoud, M., and Niederweis, M. (2007) Porins are required for uptake of phosphates by *Mycobacterium smegmatis*. *J Bacteriol* **189**: 2435–42.

Xu, M., Arulandu, A., Struck, D.K., Swanson, S., Sacchettini, J.C., and Young, R. (2005) Disulfide isomerization after membrane release of its SAR domain activates P1 lysozyme. *Science* **307**: 113–7.

Xu, M., Struck, D.K., Deaton, J., Wang, I. N., and Young, R. (2004) A signal-arrest-release sequence mediates export and control of the phage P1 endolysin. *Proc Natl Acad Sci U S A* **101**: 6415–20.

Young, R. (2002) Bacteriophage holins: deadly diversity. *J Mol Microbiol Biotechnol* **4**: 21–36.

Young, R. (2005) Phage Lysis. In *Phages: Their Role in Bacterial Pathogenesis and Biotechnology*. Waldor, M., Friedman, D., and Adhya, S. (eds). American Society for Microbiology, Washington, DC. pp. 92–127.

Young, R., and Wang, I. N. (2006) Phage lysis. In *The Bacteriophages*. Calendar, R. (ed.). Oxford University Press, New York. pp. 104–126.

Zhang, N., and Young, R. (1999) Complementation and characterization of the nested Rz and Rz1 reading frames in the genome of bacteriophage lambda. *Mol Gen Genet* **262**: 659–67.

Zuber, B., Chami, M., Houssin, C., Dubochet, J., Griffiths, G., and Daffé, M. (2008) Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. *J Bacteriol* **190**: 5672–80.